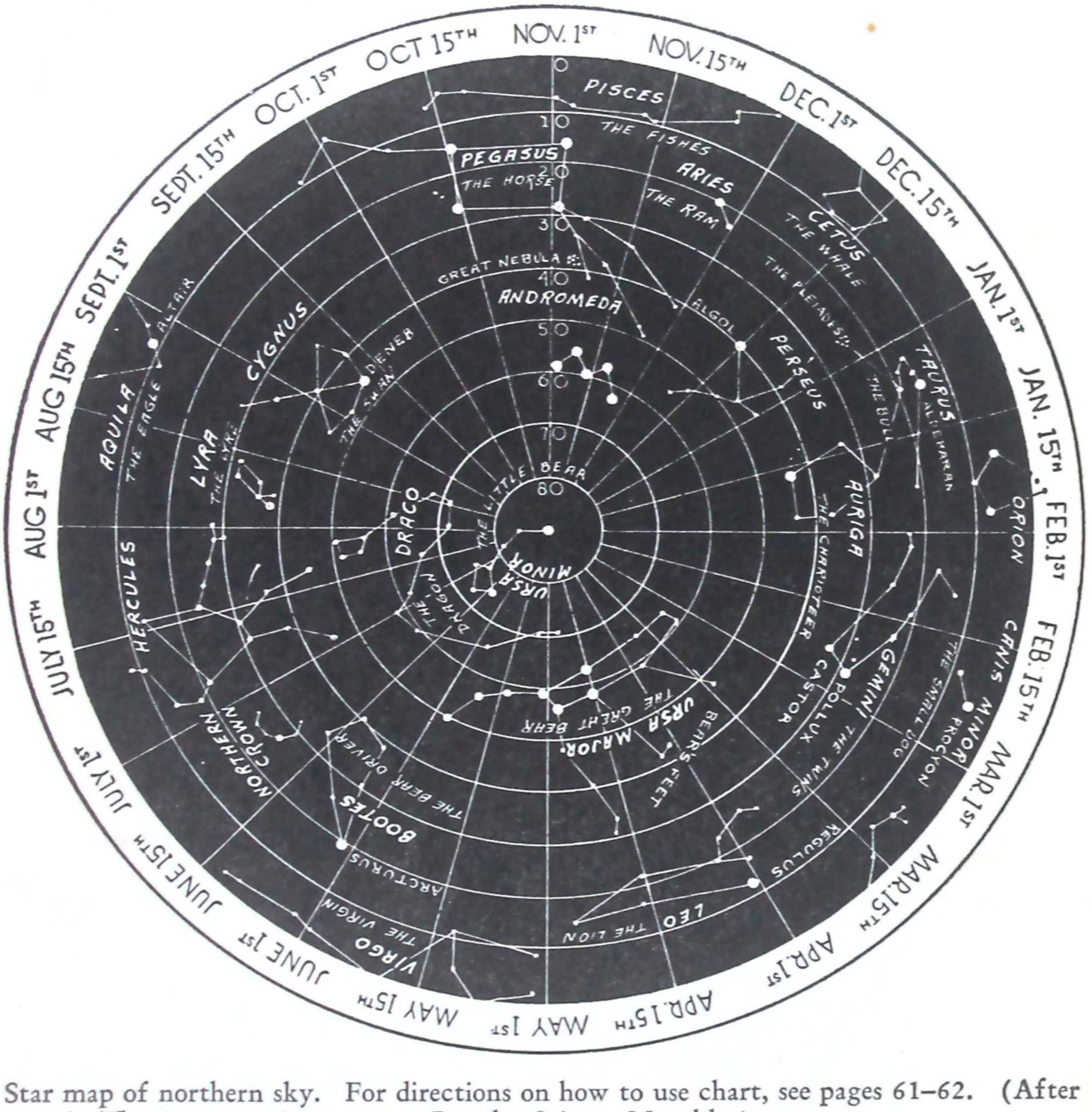
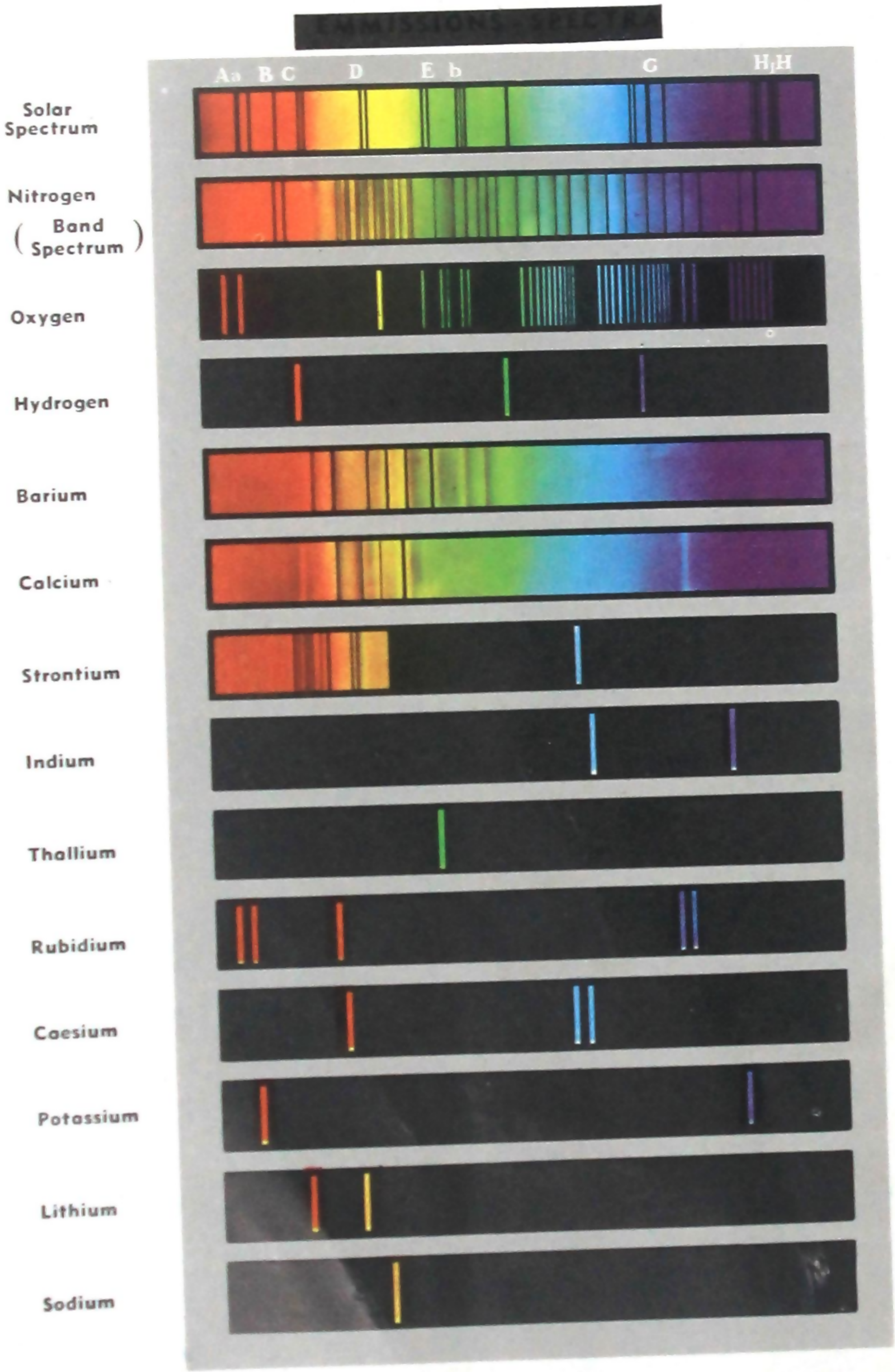


Star map of northern sky. For directions on how to use chart, see pages 61–62. (After map in The Amateur Astronomer, Popular Science Monthly.)

MAN
and his
PHYSICAL WORLD



EMISSIONS-SPECTRA



Examples of various types of spectra. Solar spectrum is dark line; nitrogen spectrum is band type; all others are bright line. (From Foster's " Inorganic Chemistry," D. Van Nostrand Co., 1936.)

MAN AND HIS PHYSICAL WORLD

DWIGHT E. GRAY

Physicist, The Applied Physics Laboratory, Johns Hopkins University
Formerly Associate Professor of Physics, The University of Akron

SECOND EDITION, TWELFTH PRINTING

D. VAN NOSTRAND COMPANY, INC.

PRINCETON, NEW JERSEY

TORONTO LONDON

NEW YORK

D. VAN NOSTRAND COMPANY, INC.
120 Alexander St., Princeton, New Jersey
257 Fourth Avenue, New York 3, New York
25 Hollinger Rd., Toronto 16, Canada
Macmillan & Co., Ltd., St. Martin's St., London, W.C. 2, England

All correspondence should be addressed to the principal office of the company at Princeton, N. J.

COPYRIGHT, 1942, 1946, BY
D. VAN NOSTRAND COMPANY, INC.

CHECKED

All rights in this book are reserved. Without written authorization from D. Van Nostrand Company, Inc., 120 Alexander Street, Princeton, New Jersey, it may not be reproduced in any form in whole or in part (except for quotation in critical articles or reviews), nor may it be used for dramatic, motion-, talking-picture, radio, television or any other similar purpose.

First Published . . . . . September, 1942

Reprinted, May, 1943; June, 1945; February, 1946; July, 1946

Second Edition . . . . . October, 1946

Reprinted, February, 1947; June, 1947; May, 1948
September, 1948; May, 1949; August, 1950; July, 1952
February, 1953; September, 1953; June, 1954
February, 1956

LAMA IQBAL LIBRARY
765

JAMMU & KASHMIR UNIVERSITY
LIBRARY
No. 22.765
Date 12.5.58
SRINAGAR

PRINTED IN THE UNITED STATES OF AMERICA

To

HELEN GRAY

Preface

This book is the outgrowth of several years of experience as the director of a science survey course required of all students except those majoring in a scientific field. Consequently the book is concerned primarily with the instruction of those students for whom it will constitute the principal, and in many cases, the only formal college contact with existing knowledge in the fields of astronomy, chemistry, geology and physics.

With a group of this sort, there can be no assumption of any extensive mathematical or scientific background on the part of the individual student. The textbook must, therefore, discuss each subject from the beginning and deliver a continued story in which each division of the subject will flow logically and naturally from the preceding one. This textbook attempts such a systematic approach in an effort to avoid creating in the mind of the student a feeling that he is being overwhelmed by an avalanche of seemingly unrelated scientific facts. The book also endeavors to maintain a close liaison between its subject matter and everyday experience and knowledge. The approach is that of the scientific method which points the way to objective observation and experiment in the solution of problems relating to natural phenomena; fundamental principles have been given precedence over new developments whenever space limitation prevents both from being treated in detail.

The text is supplemented with three types of questions: discussion questions, multiple choice questions, and true-false exercises. The discussion questions are intended to be useful in promoting classroom discussion, while the multiple choice and true-false exercises are based entirely on the textbook and may be used for grade-determining quizzes.

The specialist in any one of the physical sciences may object to the omission of various topics which he regards as essential for an understanding of the field. Nevertheless, such a lack of completeness must be an inherent characteristic of any survey book and the author must assume responsibility for the selection of topics. This has been done in the case of the present text but with no pretense of infallibility and the author will welcome any criticisms and suggestions from those who may use the book in classes.

For the present edition, various portions of the earlier text have been modified to take account of war-born technical developments. New material has been introduced in the fields of radar, reaction engines and atomic energy.

Acknowledgments

No author would care to present for use a textbook which attempts to survey a number of scientific fields without first having the manuscript read and criticized by specialists in those sciences. I feel that I have been exceedingly fortunate in having the following group of people assist me in this manner and I wish to express herewith my gratitude to them. A large part of the manuscript was read by Dr. H. P. Knauss, Monsanto Chemical Company; and by Professor F. F. Householder, Chairman of the Physics Department, University of Akron. Both offered valuable suggestions which were incorporated into the book. The chapters on Rubber and Plastics were revised and brought up to date for the present edition, following the war, by Dr. Eugene Haas, Chairman, Chemistry Department, Westminster College. Others who read and criticized chapters, in one or both editions, dealing with material in their special fields were Dr. Paul Acquarone, Department of Geology, University of Akron; Dr. Howard I. Cramer, Director of Research, Sharples Chemicals Inc.; A. G. Norris, Technical Writer, Applied Physics Laboratory, The Johns Hopkins University; J. W. Schade, formerly Director of Research, B. F. Goodrich Co.; Professor R. H. Schmidt, Department of Chemistry and Registrar, University of Akron; and Professor C. R. Upp, College of Engineering, University of Akron. In addition to these I am also deeply indebted to Dr. and Mrs. Robert Thackaberry, Department of English, University of Akron, and Miss M. Kathleen Ahern, Assistant Editor, Harvard Underwater Sound Laboratory, for checking the manuscript for grammar and English usage; to Miss Jacqueline Horst, art student, University of Akron, and Miss Phyllis Hoseit, Applied Physics Laboratory, The Johns Hopkins University, for all drawings not otherwise credited; and to publishers and industrial firms who so generously permitted me to use the quotations and illustrations credited to them in the book. To my wife, Helen Gray, I owe and wish to acknowledge formally a particularly large debt of gratitude both for her direct aid in typing and proofreading and, what is more important, for her constant encouragement and constructive criticism while the manuscript was being written.

Dwight E. Gray

Contents

Introduction

CHAPTER

1. SCIENCE AND THE SCIENTIFIC METHOD . . . . .
Definition of science. The scientific method. The purposes of science.

Man and the Universe

2. THE SOLAR SYSTEM . . . . . .
Historical background of man's astronomical knowledge. General picture of the solar system. The sun.

3. THE SOLAR SYSTEM (continued) . . . . . .
The planets. The planetoids. Comets, Meteors. The origin of the solar system.

4. THE UNIVERSE BEYOND THE SOLAR SYSTEM . . . . .
Our galaxy. Stars and constellations. Systems beyond our galaxy.

5. ASTRONOMICAL MEASUREMENTS . . . . . .
The telescope. Determination of composition, distance, size, temperature, velocity, and mass of stars and planets.

6. THE EARTH AS AN ASTRONOMICAL BODY . . . . .
Its physical characteristics. Phenomena which depend on the earth's motions and position in space. The moon and its relation to the earth.

7. THE EARTH AS AN ASTRONOMICAL BODY (continued) .
Division of the earth into centrosphere, lithosphere, hydrosphere, and atmosphere. Weather.

Man and the History of the Earth

8. FACTORS WHICH CHANGE THE SURFACE OF THE EARTH .
Beginning of earth science. Types of rock in the earth's crust. Relief-levelling factors. Relief-building factors.

9. CLUES TO EARTH HISTORY . . . . . .
The age of the earth. Rock correlations as clues to earth history. Fossil remains as clues to earth history.

CONTENTS

R PAGE

GEOLOGICAL HISTORY 169
Geological cycles and their characteristics. Geological eras and periods. Outline of major occurrences during earth history.

Man and the Tools of Science

NUMBERS AND NUMBER SYSTEMS 187
Mathematics as a language of size. Number names and bases. The writing of numbers. Features of Roman, Greek, Mayan, and Hindu-Arabic number systems.

UNITS OF MEASUREMENT 201
The role of measurement in science. Fundamental units of length, mass, and time. Calendars. Derived units of measurement.

Man and Materials

THE NATURE OF MATTER 218
Early history of the problem. Kinds of matter and their relation to each other. States of matter and some properties of each. The phenomenon of change of state.

THE NATURE OF MATTER (continued) 239
Heat and temperature. Some experimental facts relating to heat and matter. The caloric theory of heat. The kinetic theory of heat and matter.

AIR AND WATER 260
Nature, properties, and uses of air and its constituents. Nature, properties, and uses of water and its constituents.

FUELS 278
Factors influencing combustion. Solid fuels. Liquid fuels. Gaseous fuels.

METALS AND ALLOYS 295
General discussion of metals and metallurgical processes. Iron and steel. Other metals. Alloys.

COLLOIDS AND RUBBER 319
The colloidal state of matter. Rubber — its sources, properties, processing, and applications. Synthetic rubber.

SYNTHETICS 350
Definition of a synthetic. Inorganic synthetics including glass and ammonia. Organic synthetics including elastics, plastics, dyes, and drugs.

CONTENTS

Man and Energy

PTER

FORCE, WORK, AND POWER . . . . . .

Definition, kinds, and measurement of force. Definition and measurement of work. Machines as aids in doing work. Definition and measurement of power.

ENERGY AND ITS TRANSFORMATIONS . . . . . .

Relation between energy and work. Kinds of energy. The law of conservation of energy. The sun as the source of our energy.

HEAT ENGINES . . . . . .

Definition of a heat engine. Background and operation of steam engines, steam turbines, gas engines, Diesel engines, reaction engines.

MAGNETISM . . . . . .

Historical background. Experimental facts about magnetism. Molecular theory of magnetism. Types and uses of magnets.

STATIC ELECTRICITY . . . . . .

Historical background. Present picture of the nature of negative and positive electricity. Description of various static phenomena.

CURRENT ELECTRICITY . . . . . .

Relation to static electricity. Various methods of generating an electric current. The electric circuit and its characteristics. Applications involving heating effects, mechanical effects, and chemical effects of electrical currents. Transmission of electrical energy.

ELECTRICITY AND THE ATOM . . . . . .

Review of earlier discussion of nature of matter. Additional phenomena including flow of electricity through gases, discovery and investigation of the electron, photoelectric effect, and radioactivity. Atomic models. Isotopes. Atomic energy and the atomic bomb.

ELECTRONICS . . . . . .

Relation to electricity. Vacuum tubes and photoelectric cells. Applications. Nature of radar and its uses in war and peace. The proximity fuze.

WAVE MOTION . . . . . .

Wave motion as a method of energy transmission. Essentials for the production of waves. Kinds of waves. Properties of waves.

CHAPTER PAGE

29. SOUND 545
Essentials for the production of sound waves. Properties of sound waves. Sound phenomena. The physical basis of music.

30. LIGHT 572
Various light phenomena. The nature of light as indicated by the experimental facts.

31. LIGHT (continued) 598
The electromagnetic spectrum. The physical basis of seeing. Color.

Man Applies Science

32. COMMUNICATION 622
Important features of communication methods. Early development of methods of communication. Printing. The telegraph. The telephone. The radio. Other communication methods.

33. ILLUMINATION 636
General illumination problems. Advantages and disadvantages of the various sources of artificial light.

34. REFRIGERATION 657
Absorption of heat and change of state, and their relation to refrigeration. Description of various types of refrigerating devices.

READING LIST 671

INDEX 681

Introduction

Chapter 1

SCIENCE AND THE SCIENTIFIC METHOD

Definition of Science

Words need not be unfamiliar in order to be difficult to define. Many rather common ones which are constantly used in everyday conversation seem almost to defy exact definition. The word *Science* is such a term. In its broadest sense it might be thought of as embracing all knowledge, a natural interpretation in view of its derivation from the Latin verb *scire*, meaning " to know." It is a familiar fact, however, that common usage over a long period of time frequently causes a word to acquire a meaning different from that implied by its derivation. Such has been the fate of " science," which in our present day thinking has a meaning much more specialized than its origin would indicate.

As man's knowledge of natural phenomena increased, there came a time when he recognized that his growing knowledge of nature was the result of his application of a particular method of investigation. This rather well defined procedure has come to be known as the *Scientific Method*. Consequently the emphasis passed from the knowledge itself to the method by which that knowledge was obtained. From this viewpoint Loeb and Adams have defined Science [1] as " that body of knowledge which has been obtained by controlled quantitative investigation." With this definition in mind, let us explore a little further the possibilities and implications of the scientific method.

The Scientific Method

We shall approach the details of the scientific method by way of an example which consists of a problem to be solved. By analyzing a series of possible steps in the procedure of arriving at a solution to this

[1] Loeb and Adams, *The Development of Physical Thought*. New York: John Wiley & Sons, 1933, p. 2.

problem, we may be able to develop a generalization of the method. After having made the generalization we shall consider one or two other examples.

Analysis of Steps

Let us consider a hypothetical case in the field of detective fiction. We shall assume that a certain story-book detective has just discovered the body of a prominent citizen slumped over the desk in his office. Thus, *he recognizes that there is a problem to be solved*. Obviously his first step will be to get immediately as many facts as possible. A hurried investigation discloses that death was caused by gun shot and that lying near the right hand of the deceased is a revolver from which one bullet has been fired. The work has begun with the *collection of experimental facts.*

In the light of these few facts several explanations or hypotheses may suggest themselves as possibilities, such as, suicide, murder, or accident. From these our detective selects the one which at the moment seems simplest and most likely and uses it as a *tentative hypothesis.* Let us assume that the one so chosen is suicide. Next it becomes necessary to test this tentative hypothesis in the light of additional evidence, and the detective immediately begins to perform experiments directed toward this end. He looks for powder marks on the skin. He investigates to see whether or not the man was naturally right-handed. He has the bullet examined to see whether it was fired from the gun that was found beside the body. He makes inquiries concerning the existence of any motive for suicide. In other words, *he performs test experiments*, the results of which will tend either to substantiate or to contradict the suicide hypothesis. In the latter case, if the contradiction is well established, the first or suicide hypothesis will be discarded and a new one, perhaps murder, will be adopted. To be satisfactory the new one must be in agreement with both the original facts and those newly discovered. The murder hypothesis will then be tested by still more experiments.

Thus the steps in the procedure may be listed as follows:

First — The recognition of the problem.
Second — Collection of experimental facts or data.
Third — Analysis of data and setting up of a tentative hypothesis

Fourth — Performance of test experiments.

Fifth — Substantiation, modification, or abandonment of the hypothesis in the light of the results of the test experiments.

If the hypothesis is discarded as the result of the test experiments, a new one will be set up and steps three, four, and five will be repeated until an explanation is found which accounts satisfactorily for all the known experimental facts. Note that step four always provides additional experimental data. As the amount of substantiating data becomes larger and larger, the hypothesis advances to the rank of a theory and eventually may be accepted as true.

In the case of our detective story illustration, some theory will probably be proven in the last chapter to describe what actually occurred. In general, however, scientific theories can never absolutely be shown to be true, since this would require that they not only account for all the experimental facts now known, but also for all that may be discovered in the future. On the other hand, one lone, well-established fact that contradicts the theory may be sufficient to cause its overthrow. Theories concerning problems that do not permit of direct investigation, such as the nature of light and matter, or the origin of the solar system, or of life, are of this sort.

An excellent and somewhat more dignified example of the application of the scientific method exists in connection with the work of Sir Isaac Newton on the nature of light. We will discuss it in terms of the steps listed above.

First — The fundamental nature of light constituted one of a number of problems about which man had wondered and speculated for many centuries. Newton, who performed exceedingly important investigations in many fields, became interested in this problem.

Second — Newton was entirely familiar with the experimental work that had been done on light during earlier times and had added considerably to the knowledge in this field by his own experiments. He knew, for example, something about color, the laws according to which light is reflected, and the manner in which it is refracted or bent when going from one medium to another.

Third — On the basis of this knowledge he adopted the hypothesis that light consists of small particles or bits of matter which move about like tiny, high-speed bullets. Thinking of light in these terms

he was able to explain satisfactorily all the experimental facts that he knew.

Fourth — However, in explaining these data he suggested a test experiment. He had shown that if this picture of light was correct it must also be true that light would travel more rapidly through glass or water than through air. Thus, measurement of the speed of light in air and in water would either substantiate or contradict his hypothesis. Because of limitations of apparatus and experimental technique, such a measurement was not possible until a number of years after Newton's death. However, when the experiment was finally performed, it was found that light travels more slowly in water and glass than in air.

Fifth — It happened that several other experimental discoveries had led to the advancement of the wave theory of light some years before these velocity measurements were made. Consequently, the latter were not the sole cause of the overthrow of the first hypothesis and the adoption of the second. Nevertheless, they did constitute an important addition to the growing body of evidence which favored the wave picture of the nature of light over a particle interpretation. More recent discoveries have resulted in still further modification of our ideas about light. This, however, will be discussed in a later chapter.

It should be noted that in general one adopts first the most obvious hypothesis; that is, the one that at the moment seems to offer the simplest explanation of the observed facts. This hypothesis may or may not prove to be satisfactory in the light of later evidence. The scientific method is one which we use to some degree in arriving at many of our everyday conclusions. As an example, if in summer you notice a porch that has a large number of newspapers lying on it, you probably conclude that the family has gone away on a vacation and neglected to stop the subscription to the paper. In other words, you have set up a tentative hypothesis. If you were interested in investigating further, you might find that you were correct in your conclusion or you might discover a totally different explanation of the observed facts.

In coming to a conclusion about any hypothesis, the true scientist is swayed only by experimental evidence. He is not, for instance, gov-

erned principally by what he or anyone else wants the result to be, by the reputation of the man who advanced the hypothesis, by what the majority of people think about it, or by any similar emotional reaction to the problem. He will constantly check his conclusions and hypotheses by experiment and be guided solely by the results thus obtained.

Now that we have discussed the scientific method in some detail, let us look once more at the definition previously quoted. The phrase "controlled quantitative investigation" seems to imply the various steps just analyzed. "Investigation" indicates that experimental work is the basis of scientific knowledge. To be of greatest value the experiment must be "controlled," that is, the experimenter must be able to vary conditions so as to know what circumstance is responsible for any given result. To cite an instance, suppose nothing happens when you turn on the switch of a flashlight. It is probable that it needs either new batteries or a new bulb. If you replace one of these and then try the light before doing anything about the other, you have performed a controlled experiment because you thus determine just which one was at fault. An uncontrolled experiment would be to replace immediately both bulb and batteries. This would probably fix the flashlight but would not tell you where the trouble lay. "Quantitative" indicates that numerical results are to be obtained, since a science is exact only to the extent that its results can be expressed in numerical terms. Now let us look very briefly at some of the historical highlights in the growth of the scientific method.

Historical Development

A very rapid growth of scientific knowledge through the application of the foregoing method has taken place during two periods of human history. The first of these occurred principally in Greek civilization and extended approximately from 600 B.C. to 150 A.D. The second includes the time since 1500 A.D.

Before 600 B.C. — During the thousands of years of man's existence before 600 B.C., he acquired a certain amount of what might be called scientific knowledge. It was gained, however, almost entirely by trial and error methods in response to certain very practical needs. For example, as much as 8,000 or 10,000 years ago a "new stone" or

Neolithic culture probably was in existence in the valleys of the Nile and Euphrates rivers. Men of this age had no metals, but they were able to chip stones into fairly efficient tools and weapons. They were rather highly skilled in agriculture, made use of domesticated animals, and had developed some proficiency as builders and architects.

Later came the Bronze and Iron Ages, beginning somewhere about 5000 B.C. and 1500 B.C., respectively, in these same localities. With the metals which came into use during these periods, man was able to devise greatly improved tools for the advancement of his own civilization and much more efficient weapons to interfere with the development of that of his fellow men.

The Babylonians developed great skill in the study of the heavens, their accurate records serving as the basis for the later growth of Greek astronomy. More will be said in a later chapter about the remarkable records and observations of the people of Babylon and the establishment of natural units of time based on the motions of the heavenly bodies.

At no time during these centuries, however, did man get any conception of a basic orderliness in nature. He seems never to have suspected that natural events might have natural causes, causes which he might be able to discover, and having discovered, use in determining natural laws or principles. With regard to any natural phenomenon such as an eclipse of the sun or moon, or a thunderstorm or an earthquake, man's first hypothesis was always that one or more Supreme Beings controlled these events which terrified him and over which he himself had no power. Such an assumption, of course, in itself need not be contradictory to science. But it is interesting to note that man always coupled with this assumption the additional one that God did not proceed in an orderly fashion but sent these phenomena quite by whim as it might suit His passing moods of anger or kindliness. Such an assumption could never lead to the discovery of natural laws, since one began by assuming that such laws did not exist. Today we recognize that the existence of an orderly plan of operation need in no way detract from the godliness of a Supreme Being. In fact most people probably would feel that a Deity who proceeds in an orderly, lawful, cause and effect fashion is more to be respected and admired

than one subject to all the fits of anger and passion with which man himself is afflicted.

From 600 B.C. to 150 A.D. — An intellectual soil, fertile for the growth and development of a scientific attitude toward natural phenomena seems not to have existed until the Greek civilization of about 600 B.C. At that time there occurred a rather ideal balance of the various factors that would be involved in producing such a soil. For one thing, the Greek nation was wealthy and powerful, the possessor of a large slave population which performed the menial labor and left the Greeks with leisure time for study and learning. It may be noted that very seldom in history has a slave-holding nation put the leisure time thus gained to so good a use. As a matter of fact one of the greatest shortcomings of Greek science also can in part be traced to this same slave-holding factor. Since the slaves did all the manual labor, such work was considered beneath the dignity of an intellectual Greek. Consequently, theories were seldom checked by experiment, and many erroneous conclusions came to be accepted as true.

In addition to having time for the pursuit of learning, the Greeks had a good knowledge of the preceding cultures of the Egyptians and Babylonians. Likewise they were not handicapped, as some of their predecessors had been, by an organized and powerful priesthood which could set up barriers to free intellectual investigation. No paths of study or thought were closed to them by religious dogma or authority.

Among the men who made important contributions to Greek science during this period were the following, several of whom will be discussed in greater detail in later chapters:

Thales of Miletus (about 600 B.C.), was an engineer by profession but is usually thought of as the world's first mathematician. His work in this field was important because he started the science rather than because any great mathematical discoveries are credited to him, although certain propositions in geometry are probably his work.

Pythagoras (about 500 B.C.), who, following the beginnings made by Thales, laid the real foundations of the science of mathematics. His best known work was in the field of geometry where he and the school known by his name developed and arranged many of the propositions later compiled by Euclid. They also were the first to develop

to a high degree the idea of abstract numbers. The Pythagoreans believed matter to be composed of four elements, earth, air, fire, and water — a hypothesis further developed at a later time by Aristotle.

Hippocrates (about 460–370 B.C.), whose oath is still taken by most graduates of medical schools, is remembered as the father of medicine. He was particularly distinguished by his insistence on the careful collection of facts and his distrust of unverifiable theories and conclusions. Consequently, his medical methods were far in advance of any, either before his age or after it, until comparatively recent times.

Aristotle (384–322 B.C.), concerning whom it has been said that he probably influenced man's general culture more than any other individual in the history of the world. Various things contributed to his occupancy of this important position. He was fairly wealthy and had been a pupil of Plato. Therefore, he had leisure time for study as well as a good understanding of the advances made by preceding cultures. Also he had been a tutor to Alexander the Great, which gave him position and authority in the eyes of the world and made large resources of money and men available to him. At one time, for example, he had approximately a thousand men working in various parts of Asia and Greece collecting data for his natural history.

The career of Aristotle is particularly interesting because, taken as a whole, it was both a boon and a tragedy to science. To his scientific credit it may be said first, that in collecting and classifying knowledge, he was supreme in the ancient world. Although primarily a zoologist, his interests and writings dealt with every field of science and almost every kind of human endeavor. He gave the first good statement of the scientific method when he wrote, " Let us first understand the facts and then we may seek the cause," and then proceeded to pay little or no attention to it in much of his scientific work. This was particularly true in the field of physics where he seems to have done almost no checking of his speculative results against observable facts, but allowed preconceived theories and wishful thinking to have full sway. Such practices could not help hindering seriously the development of science in a world that for almost two thousand years held Aristotle's teachings to be the supreme authority on any and all questions.

Archimedes (287–212 B.C.), was the world's first great experi-

mental scientist in the modern sense of the term. He devoted most of his study to mechanics and mathematics, working out solutions to important problems connected with pulleys, levers, inclined planes, and floating bodies. Probably everyone is familiar with two stories that tradition tells about him. One of these credits him with destroying part of the Roman fleet by setting fire to the ships with sunlight focussed on them by huge curved mirrors. It almost certainly is not true. The other incident concerns his discovery, while in his bath, of the law of floating bodies — known as Archimedes' principle — and relates that in his excitement he leaped from the water and dashed through the streets of the city shouting " Eureka." Again the actual evidence for the truth of the story is not very convincing.

Ptolemy (about 150 A.D.), was the last important astronomer of the ancient world. He wrote a treatise which was a compilation of all the astronomy developed up to that time. With regard to the universe, he held a theory that thought of the earth as being at the center while the sun, moon, stars, and planets moved around it in various sorts of paths. This geocentric or " Ptolemaic " theory of the universe, which was unquestioned for fourteen centuries, will be discussed more fully in a later chapter.

From 150 A.D. to 1500 A.D. — Scientific progress in western Europe came to almost a complete standstill from about 150 A.D. to 1500 A.D. There were several contributing factors to this state of affairs. One was the worshipful veneration in which Aristotle was held by the world in general, a situation which strongly discouraged the pursuit of independent investigation. Those in authority could see no reason for experimental investigation to discover the answer to a question which could be answered much more easily by looking to see what Aristotle said about it. Closely allied with this situation was the fact that during this period the youthful Christian church attained enormous power and influence in the European world. Although it made important contributions to the society of that day from a moral and ethical standpoint, its extremely dogmatic attitude toward all natural phenomena seriously hindered the growth of scientific knowledge. Reasoning from observation and experiment, which had so increased man's knowledge of the world about him during the preceding Greek age, was largely replaced by reasoning from authority. The

dogmatism of the theological leaders of the day caused all sorts of barriers to be set up against free investigation. Such a situation, of course, could only result in stagnation as far as scientific endeavor and advance was concerned. A somewhat less important factor, although a very practical one, was the widely accepted belief that the world would come to an end about 1000 A.D., that is, at the close of the first millenium. One can readily understand that an individual who held this view might consider the scientific investigation of natural phenomena which were soon to come to an abrupt and final end a very foolish and useless sort of occupation.

The spirit of science was kept alive during this period by the Arabs who rose to considerable power in the seventh century under the leadership of Mohammed. They studied Greek science extensively and made contributions themselves to the fields of medicine, optics, chemistry, and mathematics.

From 1500 A.D. to Present. — The wane of the dominance of authority dates from about 1500 A.D. The centuries since that time comprise the second of the two periods of rapid scientific progress. One of the first men to revive the spirit of objective, unprejudiced investigation was *Leonardo da Vinci* (1452–1519). Although known most commonly as a painter and sculptor, he was also an engineer, architect, physicist, biologist, geologist, and philosopher, and was outstanding in each field. It has been said that probably no other man in the history of the world has shown so remarkable a combination of competence and versatility. A model of one of his inventions is shown in Fig. 1. In da Vinci we find a man strikingly free, for that age, from preconceptions and the influence of theological dogma and authority.

About this time, also, there began a great renaissance in astronomy. First came *Copernicus* (1473–1543) who gave the world a heliocentric or sun-centered picture of the universe. He was followed by *Tycho Brahe* (1546–1601) whose accurate observations and careful recording of data rank with the best. Overlapping to some extent the work of Brahe was that of *John Kepler* (1571–1630). Kepler is noted chiefly for his statement of three laws of planetary motion. After them came *Galileo Galilei* (1564–1642), one of the two or three outstanding scientific figures of all time. Perhaps his most important

work was his recognition of the fact that scientific advancement depends on discovering the answer to the question " How? " concerning natural phenomena, rather than on speculations and philosophizing about the " Why? " of it. Immediately following Galileo, there came that other master, *Sir Isaac Newton* (1642–1727), a man whose con-

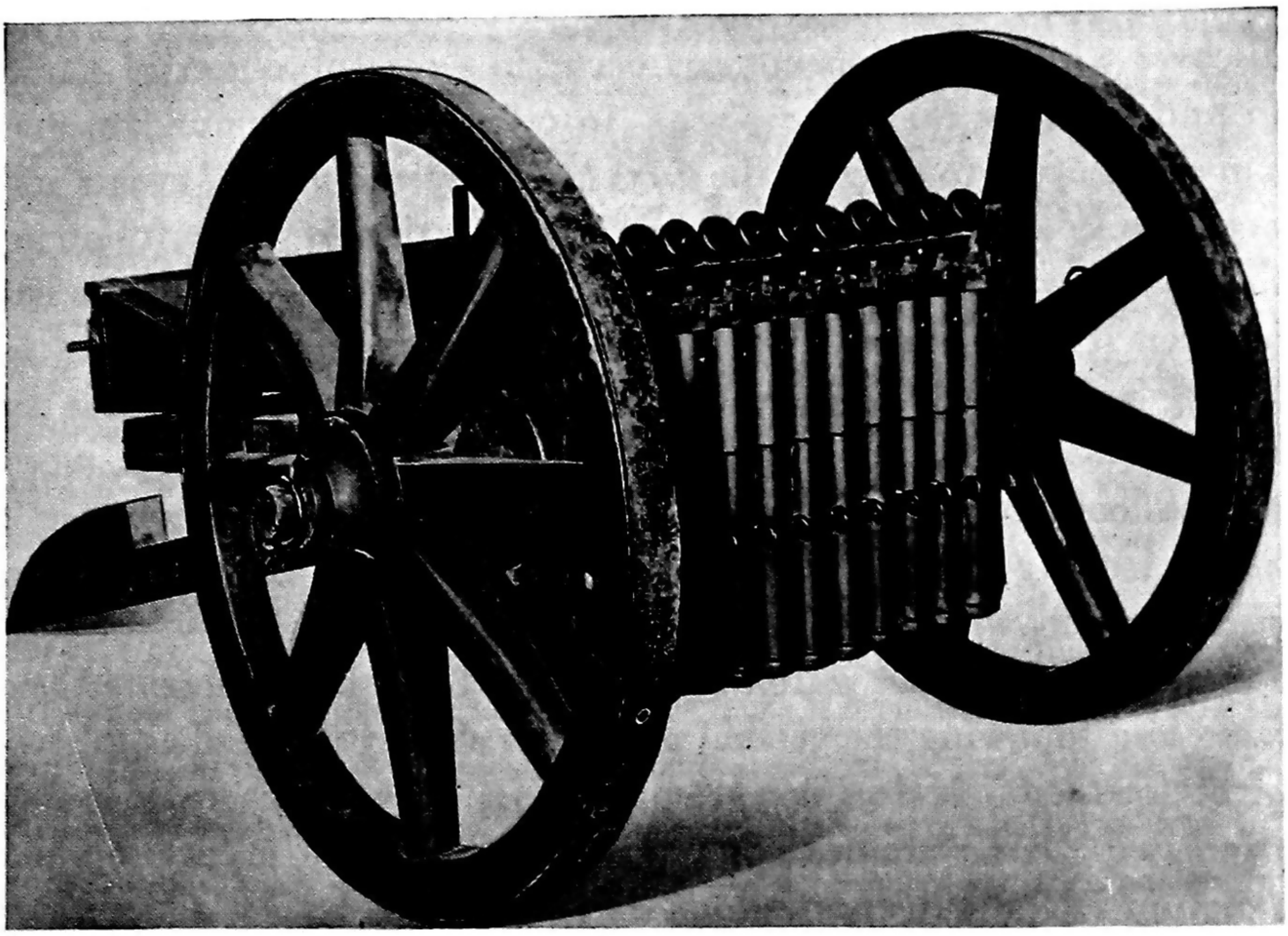


FIG. 1. Model of a multiple cannon designed by Leonardo da Vinci. Its three banks of barrels were located so that while one was firing, the second was cooling and the third was being loaded. (Courtesy Scientific American.)

tributions to any one of a number of fields of science were sufficient to place him among the world's immortals. Since the time of Galileo and Newton, there have been numerous great scientists, many of whom are fairly well known to everyone. Some of these, as well as the ones already mentioned, will be discussed at greater length in later chapters in connection with the fields in which they made their contributions.

THE PURPOSES OF SCIENCE

Before leaving the general discussion of the scientific method we may ask what we have a right to expect of science in its relation to

our natural world. We shall list two general purposes that it may be said to have.

A Description of the Facts of Experience

We have already suggested that Galileo took an important step forward when he changed the emphasis in science from attempts to discover why natural phenomena occur to a careful, objective search to find out how they take place. In other words, the fundamental aim of science is to describe the facts of nature and natural events and not, as is so commonly thought, to "explain things." Confusion on this point has probably been the cause of much of the so-called conflict between science and religion, a conflict more correctly described as having been between science and theology. The basis of science is the belief that natural events have natural causes. Consequently, when science looks for the cause of any given natural phenomenon, it is simply looking for a set of circumstances which gave rise to the event, circumstances which themselves grew out of a still earlier set of conditions. It makes this search by observing facts, by organizing these facts, and following the general procedure already outlined as the scientific method. It does not attempt to say why the chain of events originated. For example, it is within the province of science to build a picture of the universe with its myriad celestial bodies, a picture or theory which to be satisfactory must agree with all of the observed facts about the motions, positions, and other attributes of these bodies. It is not a part of the business of science to say why the sequence of occurrences was begun which led to the present astronomical organization. Such a problem involves factors which do not lend themselves to the collection of experimental data. Consequently, science should neither support nor attack any person's belief about such a question unless, of course, that belief leads him to inferences and conclusions which are in definite disagreement with phenomena that are observable.

The Formulation of Theories and Laws

A secondary purpose of science is the formulation, on the basis of experimental facts, of principles and theories which are generalizations and which will lead to new studies and increased knowledge.

There would seem to be no need to fear that we will soon have solved all the problems and made known the entire unknown. In fact, the reverse is much more probable since every new discovery seems to widen the horizon and increase the extent of our contact with unexplored areas. Scientists have not always recognized this and there have been occasions in history when they have announced that almost everything had been discovered and that the future seemed to offer little hope of further advance. For instance, in 1886 the United States Commissioner of Labor stated in his annual report that there were enough railroads, canals, international communications, and merchant shipping, and that all the world needed to do now was to settle down and enjoy itself because the next fifty years would see no such advance as the previous half century had shown. It has been pointed out that an assemblage of young people who might have read this report or heard their parents discussing it in 1886 could have included among its members Thomas Edison, thirty-nine years old; Henry Ford, twenty-three; Charles Steinmetz, twenty-one; Madame Curie, nineteen; Robert Millikan, eighteen; Orville Wright, fifteen; Guglielmo Marconi, twelve; and Albert Einstein, seven. The Compton brothers had not been born.

Today, of course, it is difficult to imagine anyone saying, " It can't be done," about almost any technical or scientific proposal. After all, we have witnessed within the past several years such astounding developments as the location by radar at night and in fog of sea and land targets, radar-guided bombing and radar-controlled gunfire, antiaircraft and artillery shells carrying five-tube radio sets in their noses, jet-propulsion engines driving airplanes at speeds approaching that of sound, and destruction of an entire city by a single atomic bomb. There can be little doubt but that man's technical intelligence rating is very high; whether he is also smart enough to benefit permanently from these advances or so stupid that he will let them destroy his civilization will probably be decided shortly.

Questions

Discussion

1. Trace the development of the meaning of the word " science," beginning with its Latin derivation.

2. Name and discuss the steps in the Scientific Method.

3. Mention one or two kinds of approaches, other than the scientific to the solution of problems and give an example of each.

4. Suggest the steps in a scientific approach to the problem of deciding:

a. Which of two candidates to vote for.

b. What make or brand of radio, automobile, tooth paste, razor blades, refrigerator, breakfast food, or other product to buy.

c. What college to attend.

d. What grocery store, drug store, filling station, or other dealer to patronize.

e. Who was to blame in an automobile accident.

f. Who should have the family car on any given night.

g. Whether the referee was partial to the opposing team.

h. Whether the administration of any given mayor, governor, or president was a success.

5. Note some of the reasons offered by people for making particular decisions on problems like the above and discuss them from the standpoint of the scientific attitude.

6. Choose one or two current advertising slogans and discuss them from the standpoint of the scientific attitude.

7. What is a " controlled " experiment?

8. Describe how you might proceed in a controlled experimental manner to find out why the automobile engine didn't start on some particular occasion when you stepped on the starter.

9. Discuss the possibility and the desirability of applying the scientific method in several fields other than those usually classed as the sciences.

10. Criticize the statement, " The purpose of science is to explain things."

11. Look up some of the contributions to human culture, other than those given in this chapter, made by Aristotle, Archimedes, and Leonardo da Vinci.

12. How can a scientific hypothesis have been of any value if it is eventually abandoned?

13. Mention one or two reasons for the lack of scientific progress during the Middle Ages.

14. Discuss the differences between the attitudes of a scientist and a propagandist toward facts that come to light in the course of an experimental investigation.

Multiple Choice

1. A scientific theory may be said to be entirely satisfactory if, (a) it accounts for over half the observed experimental facts, (b) it seems to be based on common sense, (c) almost everybody agrees with it, (d) more than half of the leading scientists in the field accept it, (e) it accounts for every bit of the experimental evidence known at the time.

2. Which one of the following would be most important in deciding whether to apply the term "scientific" to any bit of alleged knowledge? (a) The reputation of the person advancing it. (b) The method by which it was obtained. (c) The loudness and confidence with which it was stated. (d) The number of people that seemed to believe it. (e) Whether or not you wished it to be true.

3. List all of the following who performed their work during the first of the two periods of rapid scientific progress. (a) Kepler. (b) Aristotle. (c) Thales. (d) Hippocrates. (e) Galileo.

4. In the course of his work on the nature of light, Newton (a) concluded that light must consist of waves, (b) proved that light travels more rapidly through air than through water, (c) suggested a test experiment the results of which eventually established the truth of his theory of light, (d) was able to account for the way in which light is reflected by assuming it to consist of small particles, (e) completely discredited the particle theory of the nature of light.

5. List all of the following that occurred previous to 100 A.D. (a) A heliocentric theory of the universe had been suggested. (b) Four of Jupiter's eleven moons had been observed. (c) It had been proved by direct observation that the Milky Way consists of individual stars. (d) The phases of Venus had been observed. (e) A picture of the universe which put the earth at the center had been advanced.

True-False

1. Science may be said to teach man how he can safely violate the laws of nature.

2. In general, a scientific theory is supported by a larger body of experimental evidence than is a scientific hypothesis.

3. A belief that hesitancy to express a confident opinion on any and all subjects indicates weakness is characteristic of the scientific attitude.

4. The scientific attitude would be in agreement with the principle that "what you don't know won't hurt you."

5. Greek scientists of the period 600 B.C. to 100 A.D. were particularly noted for their insistence on checking all theories by experiment.

6. Under no circumstances whatever could the scientific attitude be a skeptical one.

7. The definition of science, as quoted in this chapter, would seem to suggest that experimental results obtained in specific numerical form are more to be desired than general qualitative results.

8. Ptolemy and Galileo were contemporaries.

9. Thales, Hippocrates, and Pythagoras all preceded Aristotle.

10. The particle theory of the nature of light may be said to have been quite satisfactory to Newton.

Chapter 2

THE SOLAR SYSTEM

Historical Background of Man's Astronomical Knowledge

The fascinating and often terrifying manifestations which he observed in the heavens probably were among the earliest objects of primitive man's curiosity. Thus, it is not surprising that his first steps along the road of scientific progress were taken in the field of astronomy. The laboratory was provided by nature, and the phenomena were constantly at hand.

From Earliest Times to 150 A.D.

We have already seen that it took many centuries for even a suspicion of the basic orderliness of nature to creep into the consciousness of these early peoples. Consequently, their principal scientific work for a long time consisted in keeping records of the occurrence of natural phenomena, with no thought of working from the observations toward general principles.

Somewhat later such a search to develop basic laws did get started. Often it was linked with attempts to measure time, since a fairly accurate basis for the prediction of seasons was of considerable importance to these agricultural peoples. For example, as early as 2000 B.C., the Babylonians had established a year of 360 days, and, from their observations, had learned to predict eclipses of the sun, and to chart with respect to the stars the travels of the sun, moon, and the five known planets.

When we come to the Greek civilization of later centuries, we find advances in astronomy that were remarkable in view of their lack of even the simplest kind of telescope. Around 300 B.C., Aristarchus advanced the idea that the sun might be at the center of the universe.

As we shall see, this suggestion was destined to start a bitter controversy when it came to light again some 1,800 years after Aristarchus' death. A little before 200 B.C., an astronomer named Eratosthenes measured the circumference of the earth by an astronomical method which gave a value within 3 per cent of the correct amount, although by 1492 A.D., people no longer believed the earth was of a shape to have a circumference.

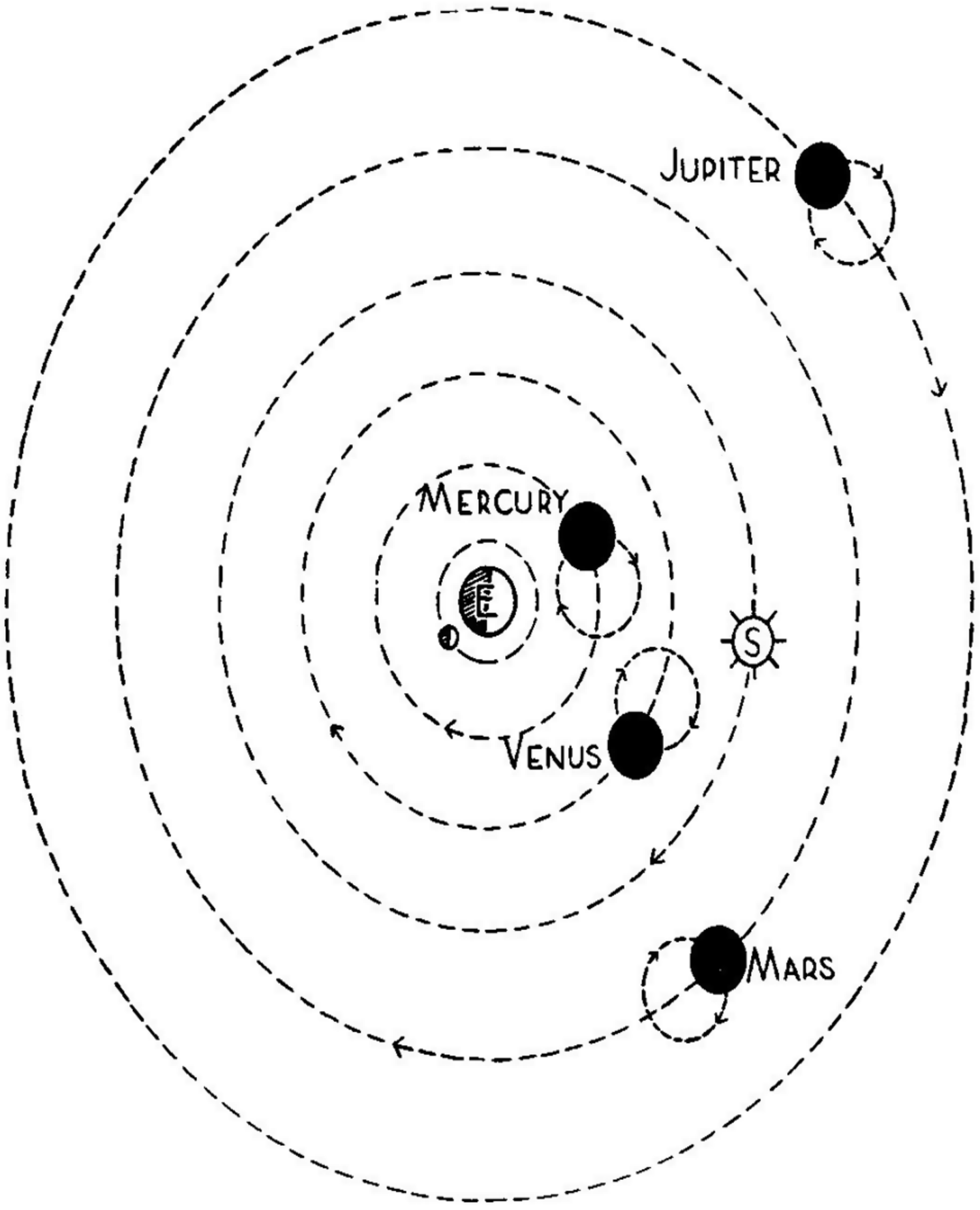


FIG. 2. Diagram illustrating Ptolemy's geocentric universe.

The two most outstanding ancient astronomers were Hipparchus and Ptolemy. The former, who worked from about 160 B.C. to 127 B.C., placed the new science of astronomy on a firm geometrical basis. He chose a theory of the universe which pictured the earth as being at the center, a hypothesis that was later enlarged by Ptolemy. Hipparchus' most astounding astronomical feat was his listing correctly in order of their brightness some 1,080 of the brightest stars in the sky. The reader will get some idea of the extent of this accomplishment if on the next clear night he will look at the sky and try to decide just how he would go about identifying for future reference.

and then arranging in order of decreasing brightness, the thousand brightest stars he can see, making all observations without the aid of a telescope.

The second of these two men, Claudius Ptolemy, is known to have worked in Alexandria from about 127 A.D. to 150 A.D. There is some evidence to indicate that he was born in Greece. Basing his ideas for the most part on the observations of Hipparchus, he advanced, around 150 A.D., the picture of the universe that came to be known as the Ptolemaic hypothesis. This model, shown in Fig. 2, pictured the earth as being spherical and located at the center of what was then called the universe but what we would call the solar system. The sun and moon were thought of as traveling about the earth in orbits in order to account for the apparent rising and setting of these bodies. Ptolemy pictured the planets as moving about the earth in the sort of complicated orbits indicated in the diagram. This was necessary in order to explain their observed behaviors. To explain the fact that the stars themselves seemed to rise and set periodically while maintaining a fixed relationship to each other, he suggested that they were embedded like gems in the inner surface of a huge sphere which made one complete rotation each day.

From 150 A.D. to 1500 A.D.

From the time of Ptolemy until about 1500 A.D., astronomy was at a standstill, as was science in general. Since, during this period, rule was almost entirely by authority, very little experimentation was carried on. Although the Ptolemaic model satisfied every experimental fact known to man at that time, its universal popularity was due chiefly to the fact that it placed man's earthly home at the center of everything, and made his planet the one about which all the others moved in respectful and majestic procession. Such a picture added immensely to man's already fairly well-developed sense of his own importance and made the hypothesis doubly hard to overthrow 1,400 years later when experimental evidence that contradicted it was discovered.

Since 1500 A.D.

In 1543, *Nicolai Copernicus* (1473–1543), a mathematics professor at Rome, published a book which carried the innocent sound-

ing title, "On the Revolutions of the Heavenly Bodies." In the light of his own observations and his studies of preceding work, he suggested that the observed phenomena involving sun, moon, planets, and stars could be explained much more simply if the sun was thought of as being located at the center of the universe. With such an arrangement the planets' observed behavior could be accounted for if they traveled in circles instead of the complicated epicycles that Ptolemy had assumed for them. The earth was to be considered simply one of the planets. He showed that a simple rotation of the earth would then explain satisfactorily the apparent rising and setting of all the other heavenly bodies. The suggestion was immediately denounced by the authorities as being anti-religious since it demoted man from his position of importance at the hub of the universe. However, as far as actual experimental evidence was concerned, either picture accounted for everything that was known at that time. The sole advantage that the Copernican or heliocentric picture enjoyed for the time being was that it required fewer and less complicated basic assumptions. Copernicus had suggested that, if his theory was correct, the planet Venus should exhibit phases as does our moon, but the performance of this test experiment had to await the invention of the telescope.

Then, in 1564, the man was born who came to be known as the father of modern science. This man was *Galileo* (1564–1642) and, as has been suggested, his work was of fundamental importance in many scientific fields. At this point we are primarily interested in his astronomical discoveries because they provided the experimental justification for the heliocentric hypothesis.

The telescope was invented about 1609 by a Dutch spectacle-maker. As soon as he heard of it, Galileo recognized its astronomical possibilities and proceeded to build one. The occasion on which he first turned it toward the sky was a significant one for science because it initiated a series of discoveries which revolutionized man's ideas about his universe.

In the first place, Galileo noted that Venus did pass through the phases predicted by Copernicus, the first actual experimental evidence favoring the heliocentric picture over the geocentric one. He saw four of the eleven moons which we now know Jupiter possesses — a

phenomenon which did not prove the correctness of Copernicus' idea, but did offer an example of the sort of motion involved in his theory. Galileo observed that the Milky Way was composed of a vast number of individual stars. He was the first man to see the sunspots which have been so publicized in recent years in connection with their possible effects on earthly weather and radio reception. By continued observation of these sunspots he detected the rotation of the sun and discovered that its period of rotation is less at the equator than at the poles, a circumstance which is part of the evidence for the gaseous nature of the sun.

Since these discoveries seemed to argue for a design of the universe of the sort advanced by Copernicus, bitter controversy with established authority immediately arose. Galileo's fellow professors at the University of Pisa flatly refused to look through his telescope lest they see something they already knew could not exist. An outstanding theologian of the day preached a sermon on the paraphrased Biblical text, " Ye men of Galileo; why stand ye here idly gazing up into heaven? " The entire heliocentric idea was denounced as being anti-religious and contrary to the teaching of the Bible. Finally, Galileo was brought before the Inquisition and forced publicly to recant upon penalty of death. By saying for publication that he had been wrong and that the earth really did not go around the sun, he saved himself from torture and death, although he was kept a virtual prisoner the rest of his life. Meanwhile, the earth, apparently quite ignorant of the battle being waged to keep it at the top of the astronomical aristocracy, calmly continued to pursue the same path around the sun that we believe it has been following for countless ages. It is worth noting that, bitter as was the controversy at the time, after subsequent discoveries had made it evident that Copernicus had been essentially right and Ptolemy wrong, theologians and astronomers alike found it quite possible to interpret the Bible in such a manner that no conflict existed. Thus today there is probably almost no one in the civilized world who finds it impossible to reconcile his particular religious beliefs with the idea that the earth travels about the sun instead of the sun about the earth.

A contemporary of Galileo's was *Johann Kepler* (1571–1630), a

resident of Wurtemburg. Kepler enunciated three principles, based on extensive experimental evidence, that describe accurately certain relationships which concern the motions and positions of the planets in the solar system.

The first of these principles, known as Kepler's laws of planetary motion, merely states that:

I. The planets travel in paths which are ellipses, with the sun at one focus.

The second law expands the description one step further, stating something quantitative as to just how the planets proceed around their respective courses, viz.:

II. Every planet travels in its orbit so that a line drawn from the sun to the planet sweeps out equal areas in equal intervals of time.

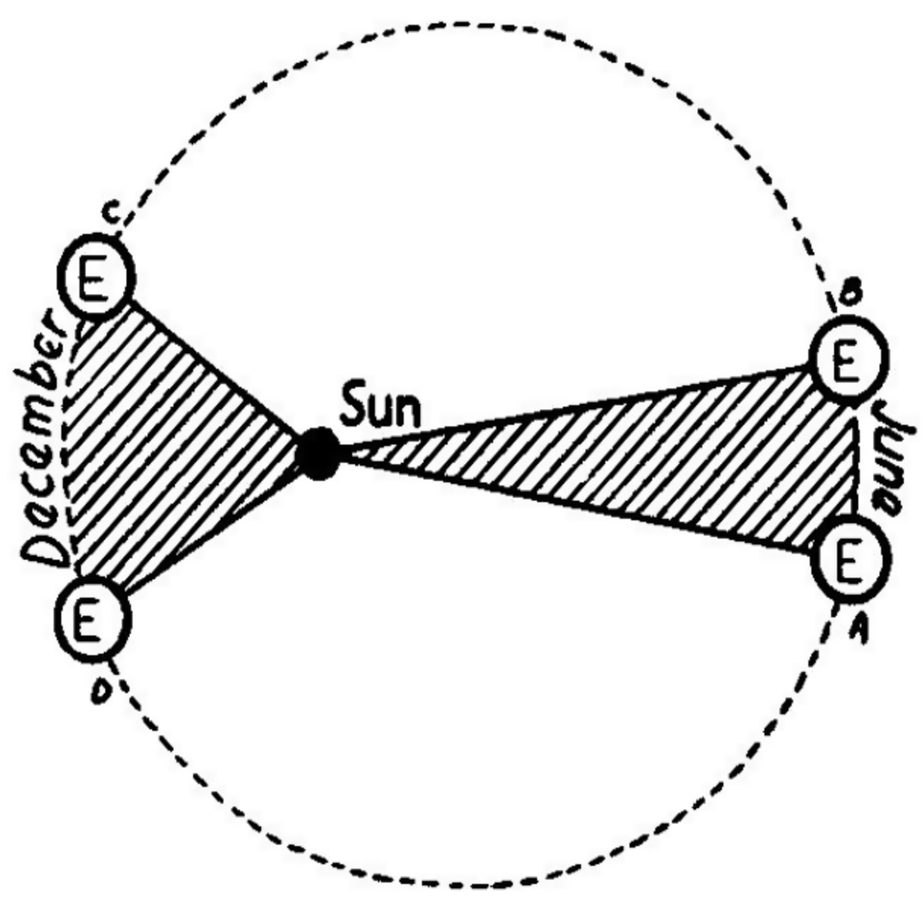


FIG. 3. Diagram illustrating Kepler's second law of planetary motion.

As will be evident from the diagram (Fig. 3), this can be true only if the planet travels with a greater velocity when it is near the sun than when it is far away. The third law relates the periods of revolution of the planets in their orbits to their relative distances from the sun:

III. The squares of the periods of revolution of the planets in their orbits are proportional to the cubes of their distances from the sun.

The year that Galileo died, *Sir Isaac Newton* (1642–1727) was born. His accomplishments, like Galileo's, were of fundamental importance in many fields. The work having the greatest bearing on our immediate discussion was his enunciation of the Law of Universal Gravitation. Kepler's laws were experimental, that is, they were exact statements of the way in which certain natural phenomena actually were observed to occur. Newton was able to show that these experimental facts are exactly what one would expect if a certain force exists between objects, a force that depends on the masses of the objects

and on how far apart they are. Therefore, he postulated the existence of such a force and described it in the statement of the law of universal gravitation, as follows:

"Every two objects in the universe attract each other with a force which is directly proportional to the product of their masses and inversely proportional to the square of their separation."

This means, for example, that the force of attraction between two one-pound objects one foot apart is half as great as that between a one-pound and a two-pound object also one foot apart; or that the attraction between two objects one foot apart is four times as great as that between the same two objects if they are placed two feet apart.

It should be noted that Newton first developed the law mathematically as a principle which would make Kepler's experimental laws seem reasonable. Actual experimental verification, which will be discussed later, came with use of the law as a basis for predicting successfully various undiscovered natural phenomena. Note also that the law makes no attempt to state why objects attract each other, only how and according to what principles they do so.

Since the day of Newton there have been many outstanding contributors to man's knowledge in the field of astronomy. The men we have mentioned made their contributions in the early part of the work and were of immense importance in putting science on the right track.

Having looked briefly at some of the persons involved in the early growth of our astronomical knowledge, let us now consider the knowledge itself, beginning with a general view of the solar system.

A General Picture of the Solar System

Before we meet the individual members of the solar family it might be well for us to take a hurried look at the group as a whole. Since our earth is a part of the solar system we are not in the best position for observation, so let us ask how we might expect it to look to an observer situated entirely outside this little group of planets. We will assume our observer to be located on the nearest celestial body to the solar system that is visible to the naked eye, a star which we call Alpha Centauri. Although it is the nearest star, it still is not exactly next door as we measure earthly distances. As a matter of fact to go

from the Earth to Alpha Centauri would require a trip of some 26 trillion miles. Thus it is evident that distances measured in miles from the solar system to some of the more distant stars would involve very large and very awkward figures, much worse, for example, than if one had to express the length of an automobile or airplane trip in inches. To avoid these large numbers, astronomers express stellar distances in terms of a unit called the *Light Year*. This is defined as the distance light would travel in one year and is equal to about six trillion miles.[1] On this basis Alpha Centauri and our hypothetical observer are distant a matter of some 4.3 light years. Now if his powers of vision were about like our own and if he were equipped with a telescope no better than our best one, he would not be greatly impressed with us because, of our entire solar system, the sun is all that he would be able to see. This he would see as a bright star, but not the brightest in his sky. The planets he would not see at all, which suggests immediately that Alpha Centauri and any or all of the other stars we can see on a clear night may have families of planets circling them, since such bodies would be quite beyond our powers of vision. However, if we equip our hypothetical Centaurian observer with a super telescope and endow him with vision far superior to our own, we believe he would get somewhat the following general impression of this small collection of heavenly bodies we call the solar system. Its principal members are shown in the diagram in Fig. 4.

First, as has been indicated, he would see the sun as a first magnitude star. Traveling in paths about this star, he would notice at least nine other bodies and close inspection might show him that the orbits of these bodies are slightly elliptical in shape. He would note that these nine planets seem to be divided into an inner group of four and an outer group of five with a relatively large separation between the orbits of the two. The inner planets, which we modestly call the earthlike, are Mercury, Venus, Earth, and Mars, in order of distance from the sun. The outer set, also in order of distance from the sun, consists of Jupiter, Saturn, Uranus, Neptune, and Pluto. All of these except Pluto are considerably larger than any of the earthlike planets. The orbits of all would appear to be in about the same plane, with

[1] Since light travels some 186,000 miles per second, 1 light year equals 186,000 × 60 × 60 × 24 × 365 = about 6,000,000,000,000 miles.

Pluto's tilted the most. Certain other bodies would be seen to be traveling about the planets, bodies which we call moons or satellites. Jupiter, the most favored from the standpoint of moonlight, has some eleven such satellites, while neither Mercury nor Venus seems to have any. The earth, as we know, has one satellite.

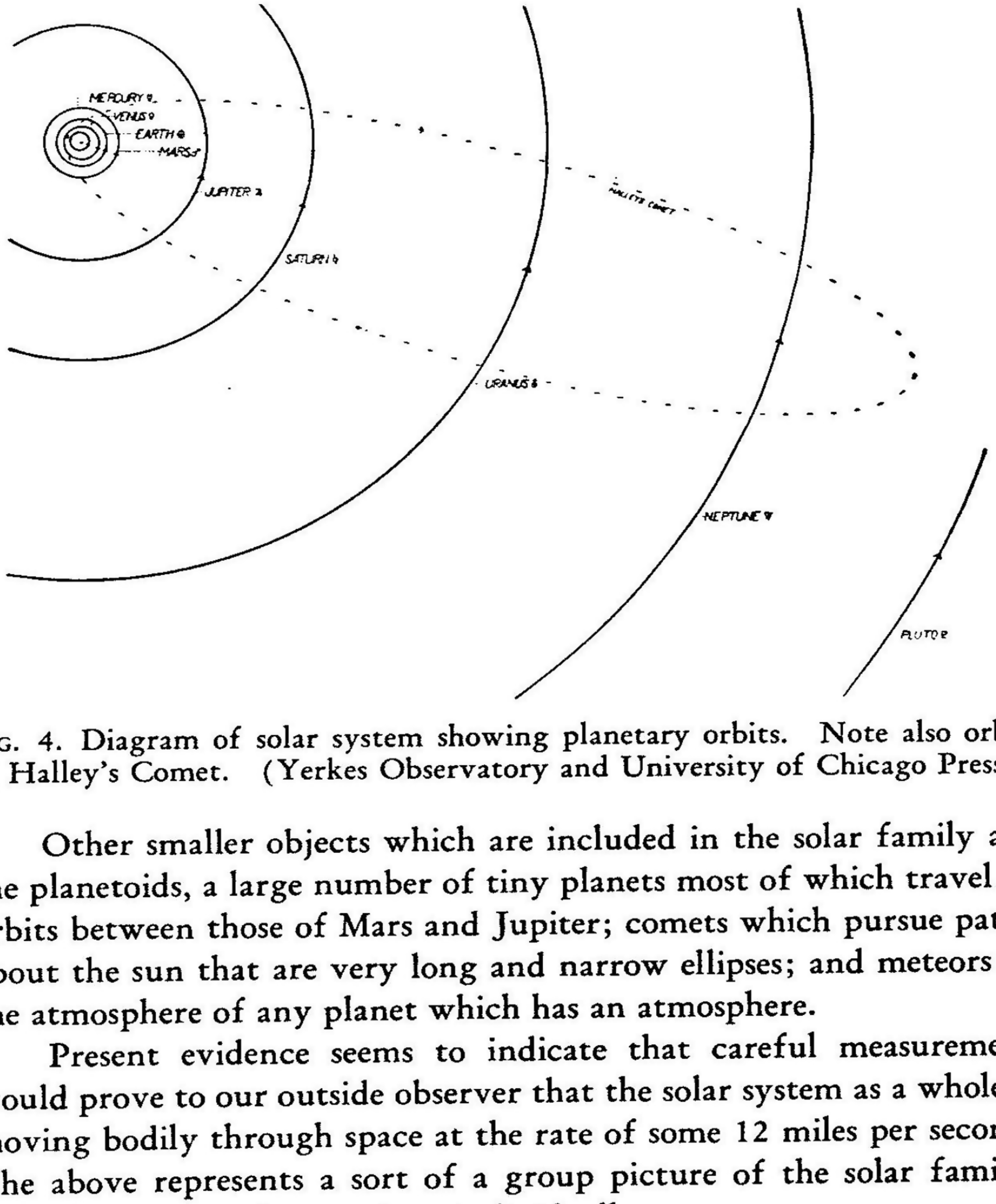


FIG. 4. Diagram of solar system showing planetary orbits. Note also orbit of Halley's Comet. (Yerkes Observatory and University of Chicago Press.)

Other smaller objects which are included in the solar family are the planetoids, a large number of tiny planets most of which travel in orbits between those of Mars and Jupiter; comets which pursue paths about the sun that are very long and narrow ellipses; and meteors in the atmosphere of any planet which has an atmosphere.

Present evidence seems to indicate that careful measurement would prove to our outside observer that the solar system as a whole is moving bodily through space at the rate of some 12 miles per second. The above represents a sort of a group picture of the solar family. Now let us meet the members individually.

Members of the Solar Family

In the solar family there is but one star — the sun. That is to say, the sun is the only member which shines continuously of its own light. This is due to its state of incandescence caused by its high temperature. The planets, planetoids, and satellites are all cold, dead bodies which are visible only because they reflect the light of the sun. In other words, if the sun were suddenly to go out, planets would not be visible to us any more than the furniture in a room at night is visible if all the lights are extinguished.

The Sun

Physical Characteristics. — The sun is far from being the most important ball of fire in space, either with respect to size or brightness. Betelgeuse has a volume some 27 million times that of the sun. There are stars so large that the earth could pursue its present orbit entirely within them and not come near the surface. On the other hand, many stars are much smaller than the sun. Similarly there are stars ten thousand times as bright and stars one ten-thousandth as bright.

The sun has a diameter of some 864,000 miles — roughly one hundred times that of the earth and ten times that of the largest planet Jupiter. Or in other words, if the diameter of the sun were to be set equal to the height of the Empire State Building, the distance through the earth would be some 11 feet. The sun's mass is some 700 times that of all the planets put together. It, as well as all the other bodies in the universe, seem to be composed of the same sort of materials we find on the earth. Of the 92 or 93 elements or building blocks of nature which we have discovered, some 61 have been found on the sun, and none have been found in the sun which are not now known on earth. If the reader is not already familiar with how such things are determined and has a proper scientific attitude, he should be skeptical of the last few statements, and wonder how anyone has a right to make them about objects so far away. Such questions we will attempt to answer in Chapter 5 on " Astronomical Measurements."

The sun exhibits a rotational motion similar to that of the earth on its axis, with the interesting difference that the rate of rotation is not the same all over its surface. On the earth, a day is 24 hours long

regardless of whether we measure it at New York, New Orleans, Rio de Janeiro, or Little America. But on the sun a point on the equator makes one complete rotation in about 25 of our days, while a day at either pole lasts for some 34 of ours. This fact was first observed by Galileo and constitutes one of the evidences from which we deduce that the sun must be gaseous. In addition to its rotation about an axis which is inclined seven degrees to the plane of the earth's orbit, the sun and the entire solar system with it is believed to be moving through space toward the constellation Hercules at a rate of some 12 miles per second.

For purposes of description and discussion, the sun has been divided into four parts or spheres. The central or main portion is called the *photosphere*. It is believed to be gaseous but of high density because of the enormous pressures to which it is subjected. The central part might be thought of as somewhat analogous to the air in a highly inflated automobile tire. It is a gas but because of the pressure has a high density and exhibits certain properties of hardness usually associated with solids. No direct evidence is available, of course, as to the temperature of the interior of the photosphere, but it has been variously estimated to be from 20 to 40 million degrees Fahrenheit. Most of the light which comes to us from the sun comes from the surface of the photosphere where temperatures have been measured to be of the order of 11,000 degrees Fahrenheit or 6,000 degrees Centigrade.

Just outside the photosphere is a layer known as the *reversing layer*. It is some 1,000 miles thick and has a lower density and temperature than the inner part. This layer becomes important in connection with the sort of spectrum emitted by the sun which will be mentioned later.

Surrounding the reversing layer and extending to a height of about 5,000 miles above it is the *chromosphere* or color sphere. Its density is very low. It contains a high percentage of hydrogen and helium, the latter having been discovered in the sun's chromosphere more than 29 years before it was found on the earth. Great bursts of colored flame called prominences shoot out from the chromosphere to distances of many thousands of miles. A photograph of one of these is shown in Fig. 5. Note the comparative size of the earth.

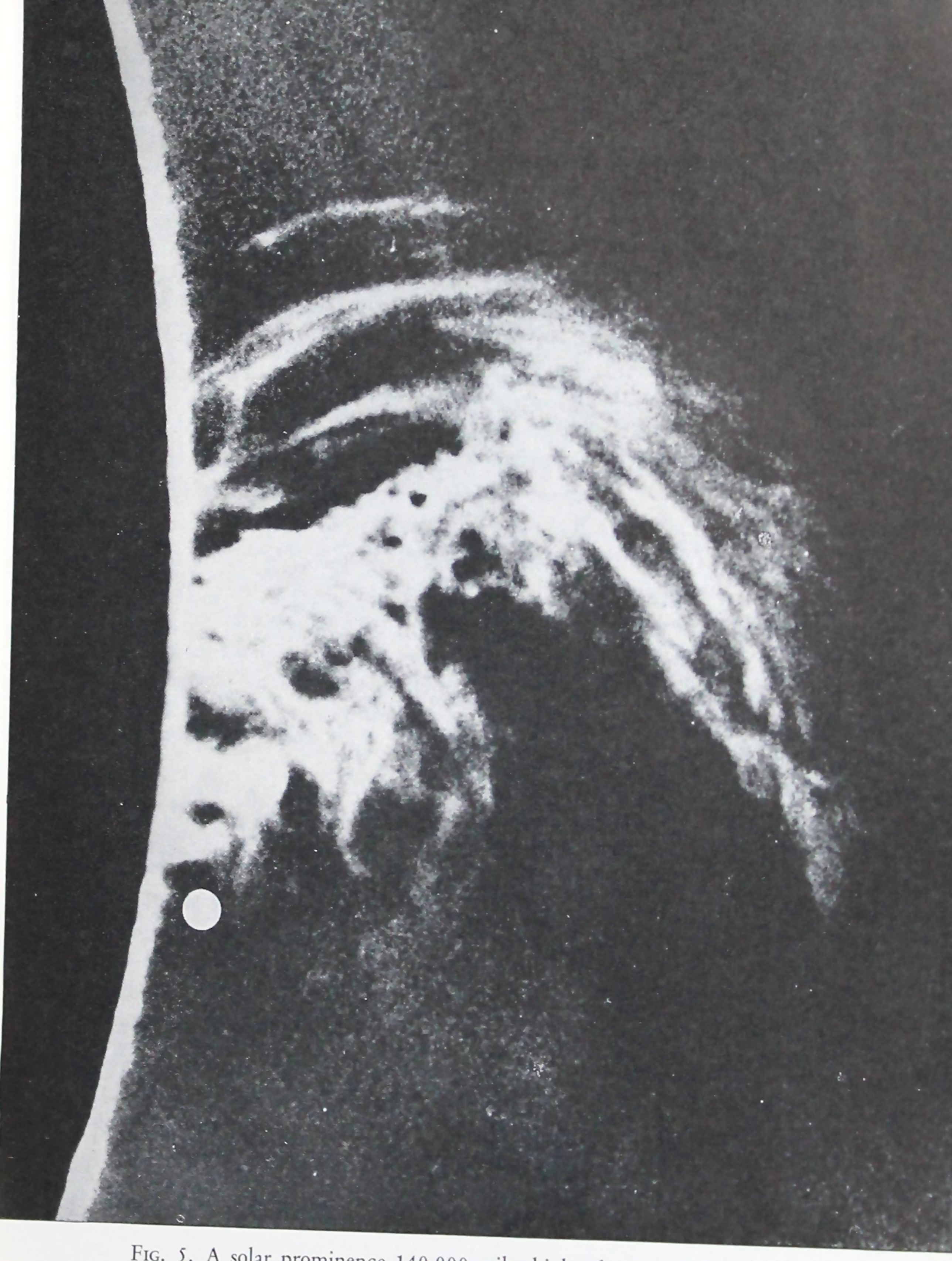


FIG. 5. A solar prominence 140,000 miles high, photographed July 9, 1917. Small white disk indicates relative size of the earth. (Mt. Wilson Observatory.)

Above the chromosphere is the *corona*, a gaseous layer whose outer boundary is at least several hundred thousand miles outside the surface of the photosphere. The corona can be studied only during a solar eclipse, when the dazzling central disk of the sun is hidden by the moon. Such investigations make it worth while for scientific expeditions to travel thousands of miles in order to be in the path of totality during an eclipse of the sun. A photograph of such a corona taken July 8, 1937, is shown in Fig. 41 on page 106.

Sunspots. — The phenomenon of sunspots provides one of the most interesting problems in connection with the study of this star. Although highly publicized during recent years, they do not represent a modern discovery. Among the first things that Galileo saw when he turned his telescope toward the sun were these dark patches or spots. In fact, there is evidence that a few very large ones visible to the naked eye had been noticed long before Galileo's time.

We believe that sunspots are huge whirlpools or cyclonic storms in the atmosphere of the sun, whirlpools varying in diameter from a few hundred miles to perhaps 100,000 miles or more. They appear dark to us because they are a thousand or two degrees cooler than the surrounding surface of the sun — the way a lighted candle might look dark if viewed against a background of an extremely bright electric light. Fig. 6 shows the largest group of sunspots ever photographed.

Much has been written in recent years and many speculations indulged in concerning the possible effects of sunspots on man and his world. From records of sunspots, kept over the past 300 years or more, it is known that the number visible varies cyclically, the average time from one maximum to the next being a little over eleven years. During the present century, maxima have occurred in 1906, 1917, 1928, and 1939. Certain earthly phenomena of a cyclic nature have been shown definitely to accompany the rise and fall of sunspots. Among these are variations in the earth's magnetic field, character of radio reception, and brilliancy of auroral displays. Somewhat less well established is the hypothesis that earthly weather trends accompany and are caused by sunspot cycles. At the same time, the studies of Professor Douglas, of the University of Arizona, in which he determined the variations in favorable and unfavorable growing weather during past centuries by examining the characteristics of tree rings,

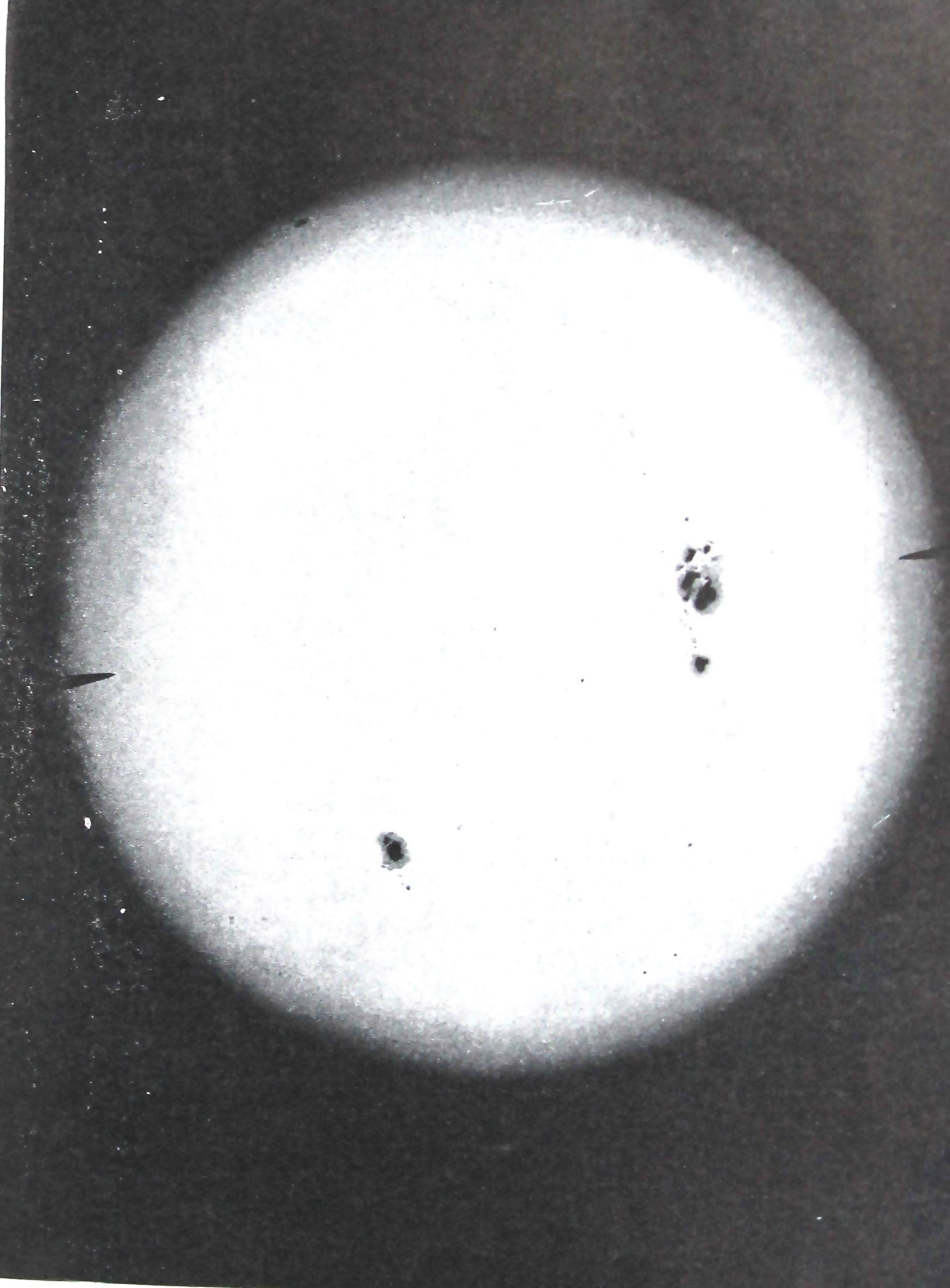


FIG. 6. Direct photograph of sun showing largest group of sunspots ever photographed. Taken January 24, 1926. Pointed markers indicate sun's axis of rotation.

show some remarkable correlations between these two. A striking example of agreement came when, some time after determining from his tree ring data that no 11-year growing cycles occurred between 1645 and 1715, he discovered that old astronomical records indicated a dearth of sunspots during the same period.

Solar Energy. — Finally, we must mention the one problem about the sun which is perhaps the most puzzling of all; that which concerns the source of the sun's energy. It is both interesting and important to us, since when we trace the ancestry of practically any kind of energy known on earth, we find its earliest ancestor to have been solar energy.

The earth's surface receives energy from the sun at an enormous rate, a rate which someone has calculated to be the equivalent of one-fifth the capacity of Muscle Shoals for every human being on earth. But only about half of the energy that strikes our atmosphere ever reaches the earth's surface, the rest being reflected and absorbed by the air. Then one must also remember that, compared with the sun, the earth is no more than a speck far out in space, so that of the total energy radiated by the sun, our earth intercepts only a small fraction, or about one- two-billionth. Evidently then, the sun is emitting energy at an inconceivably high rate. Further, geological evidence tells us that this has been going on for at least two billion years and probably much longer, all of which adds up to something of a problem when one tries to account for all this energy.

All of the earlier hypotheses were quite unable to account for such great quantities. The simplest and first suggestion was merely that the sun was composed of highly combustible materials which were burning and emitting heat. But the best fuels imaginable, in quantities the size of the sun, could not possibly account for more than a few hundred or thousand years of such prodigal expenditure of energy. Another hypothesis supposed that the sun had once been larger and hotter than at present and suggested that as it cooled and contracted energy was emitted. Again the rate of emission and the time involved were the stumbling blocks, because they required the sun originally to have been of an absurdly large size.

With modern research into the atom has come evidence to indicate that under certain conditions matter itself may be transformed

into energy. Theoretical calculations of Einstein indicate that a gram of matter, regardless of composition, would be the equivalent of 2.15×10^{13} calories of heat energy. On this basis, the energy produced by the annihilation of a buffalo nickel would be sufficient to change some 150,000 pounds of ice into steam. Such an explanation appears to account adequately for the vast quantities of energy involved and also is in line with the results of modern experiments on atomic energy. All of these hypotheses imply, of course, that the sun is being used up or is running down and will in some billions or trillions of years become a cold, dead derelict in space. This, however, is a catastrophe that is of rather greater academic than practical interest to us as inhabitants of the earth and baskers in the warm light of the stellar member of our solar family.

Questions

Discussion

1. Discuss briefly the contributions to astronomy made by Aristarchus, Hipparchus, Democritus, and Eratosthenes.

2. How did Ptolemy and Copernicus respectively account for each of the following observed phenomena? (a) Day and night. (b) The apparent rising and setting of the stars. (c) The apparent rising and setting of the moon.

3. From a scientific standpoint, how could the Ptolemaic and Copernican hypotheses have been almost equally satisfactory in 1600 and the former quite unsatisfactory in 1700?

4. How would a picture of the solar system that we might draw today differ from one drawn by Copernicus?

5. What part did the observations of Galileo play in the controversy over the nature of the universe?

6. Science and the Church bitterly disagreed about the nature of the solar system at one time and then a few years later agreed about it; can you name any other similar discussions?

7. Explain the answer that you would give to the first Multiple Choice question on this chapter.

8. State Kepler's three laws of planetary motion and give reasons for their being called laws rather than theories.

9. What is the connection between Newton's Law of Universal Gravitation and Kepler's laws of planetary motion?

10. How many times would one have to circle the earth at the equator in order to travel one light year?

11. Name the planets in order of their distance from the sun.

12. What is meant by the statement that the orbits of all the planets are in about the same plane?

13. Distinguish between the physical characteristics of stars and planets.

14. If the earth rotated in the same manner as does the sun how would the length of a day in New Orleans compare with that of one in Detroit? Explain.

15. Name the four parts or " spheres " of the sun and give the outstanding characteristics of each.

16. State and discuss the evidence for and against the various theories that have been advanced to account for the sun's energy.

17. Mention several ways in which sunspots are believed to affect the earth and state the evidence for each.

Multiple Choice

1. Which one of the following is a necessary consequence of Kepler's laws of planetary motion? (a) The planets increase in size in order of their distance from the sun. (b) Mercury's day is the same length as its year. (c) The bigger a planet is, the longer is its year. (d) A planet twice as far from the sun as the earth would have a year twice as long as ours. (e) The earth is traveling in its orbit at a different speed in January than in July.

2. The path of Pluto around the sun is (a) circular, (b) elliptical, (c) parabolic, (d) hyperbolic, (e) a straight line.

3. Which one of the following does not belong with the rest? (a) Alpha Centauri. (b) Jupiter. (c) Halley's comet. (d) The moon. (e) The largest of the planetoids.

4. List all of the following that would be in agreement with the picture of the universe advanced by Copernicus. (a) Venus exhibits the phenomenon of phases. (b) The earth travels about the sun in a circular path. (c) The earth rotates on its axis. (d) The sun travels more miles in space in the course of a year than does the earth. (e) The sun is located at the center of the universe.

5. List all of the answers in Ques. 4 which we believe today to be true.

True–False

1. All elements known on earth have also been found to be present on the sun.

2. Previous to the astronomical work of Galileo, no actual experimental evidence existed which favored the Copernican universe over the Ptolemaic.

3. The solar system contains only one star.

4. If Alpha Centauri had suddenly been extinguished 5 years ago today, we would not become aware of it for several more months.

5. The earth receives about half the energy emitted by the sun.

6. If Copernicus' picture of the universe and Kepler's second law were

both exactly correct, Mars would travel at the same speed in its orbit at all times.

7. Long before the time of Galileo, the four largest moons of Jupiter had been observed although their significance had not been recognized.

8. A point on the sun's equator rotates more rapidly than does one near either pole.

9. Certain observations of Venus made by Galileo provided evidence favoring the Copernican universe over the Ptolemaic.

10. A light year may be defined as the time required for light to travel a distance of about six trillion miles.

Chapter 3

THE SOLAR SYSTEM (Continued)

The Planets

THE fact that all the planets are cold, dead bodies which are visible to us only because they reflect sunlight has already been mentioned. The accompanying table (Table 1) lists their diameters and distances from the sun in miles together with certain other data about them. A scale model of the solar system may present a still better concept of the relationships of the various planets to each other.

A Scale Model of the Solar System. — In describing our model we shall enlarge on the one already presented which pictured the sun as having a diameter equal to the height of the Empire State Building, a matter of some 1,248 feet or 102 stories. Let us also assume the sun to be located at the site of this skyscraper. On such a basis then, Mercury might be visualized as a four-foot meteorological balloon floating over the airport in Newark, New Jersey. An 11-foot sphere of polished aluminum located along the Hudson River a few miles below Tarrytown, New York could represent the glamorous Venus. The earth, a little larger than Venus but not nearly as beautiful, would also lie along the Hudson, a few miles above Tarrytown. To Peekskill, New York, would go the privilege of possessing the reddish six-foot sphere that is Mars. The present state capital building in Albany might be replaced by a globular 12-story structure that would represent Jupiter. Saturn, a spherical 10-story building in Washington, D. C., would have around its equator an ill-proportioned porch some 76 feet wide and less than half an inch thick, if its rings were to be included in our model. Toronto, Ontario, might point with pride to Uranus, a 44-foot sphere, while Neptune, a few feet larger, would be one of the principal points of interest to tourists in Indianapolis, Indiana. Finally, distant Pluto, probably about the size of Mars or smaller,

	Mercury	Venus	Earth	Mars	Jupiter	Saturn	Uranus	Neptune	Pluto
Discovery	Ancient (Speed)	Ancient (Beauty)		Ancient (War)	Ancient (King of Gods)	Ancient (Time)	Herschel 1781	Galle 1846	Tombaugh 1930
Distance to Sun in Millions of Miles	36	67	93	142	483	886	1782	2790	3700
Diameter in Thousands of Miles	3	7.7	7.9	4.3	88	74	31	35	Probably 4
Period of Rotation	2.9 Months	Unknown	1 Day	1 Day 44.5 Min.	9 Hr. 50 Min.	10 Hr.	11 Hr.	16 Hr.	Unknown
Period of Revolution	2.9 Months	7.4 Months	12 Months	22.6 Months	12 Years	30 Years	84 Years	165 Years	249 Years
Appearance	Observation Difficult	Very White		Reddish Canals (?)	Pronounced Cloud Bands	Yellow to Green	Green		
Atmosphere	None	Extensive	Extensive	Less than Earth	Permanent Clouds				
Temperatures	Melting Lead to Liquid Air	Unknown Clouds		Probably Low	— 200° F. on Surface	— 300° F. Average	Very Low	Very Low	Very Low
Satellites			1	2	11	9	4	1	
Mass (Earth = 1)	.05?	.8	1	.11	314	94	14	16	

might turn up in a park in St. Louis, Missouri. If we wished to include the star Alpha Centauri in our model and measured from New York in finding its location, we would discover that some 260 complete trips around the world would be required to represent the 4.3 light years that separate it from the solar system. Although the important data concerning the nine planets are given in Table 1, a few of the striking features of each should be mentioned.

Mercury, the smallest planet and the one nearest the sun, is also the fastest-moving and so is aptly named for the messenger of the gods in Greek mythology. The lack of observable markings on its surface make the determination of this planet's day difficult. Its period of rotation is believed to be the same as its period of revolution about the sun. If this is true it means that Mercury always keeps the same face toward the sun, putting a little over half of its surface in perpetual sunlight where temperatures may average as high as that of molten lead, and the balance in everlasting darkness at temperatures far colder than the solid carbon dioxide that we know as "dry ice." Mercury exhibits to us the phenomenon of phases which we are most familiar with in the case of our own moon.

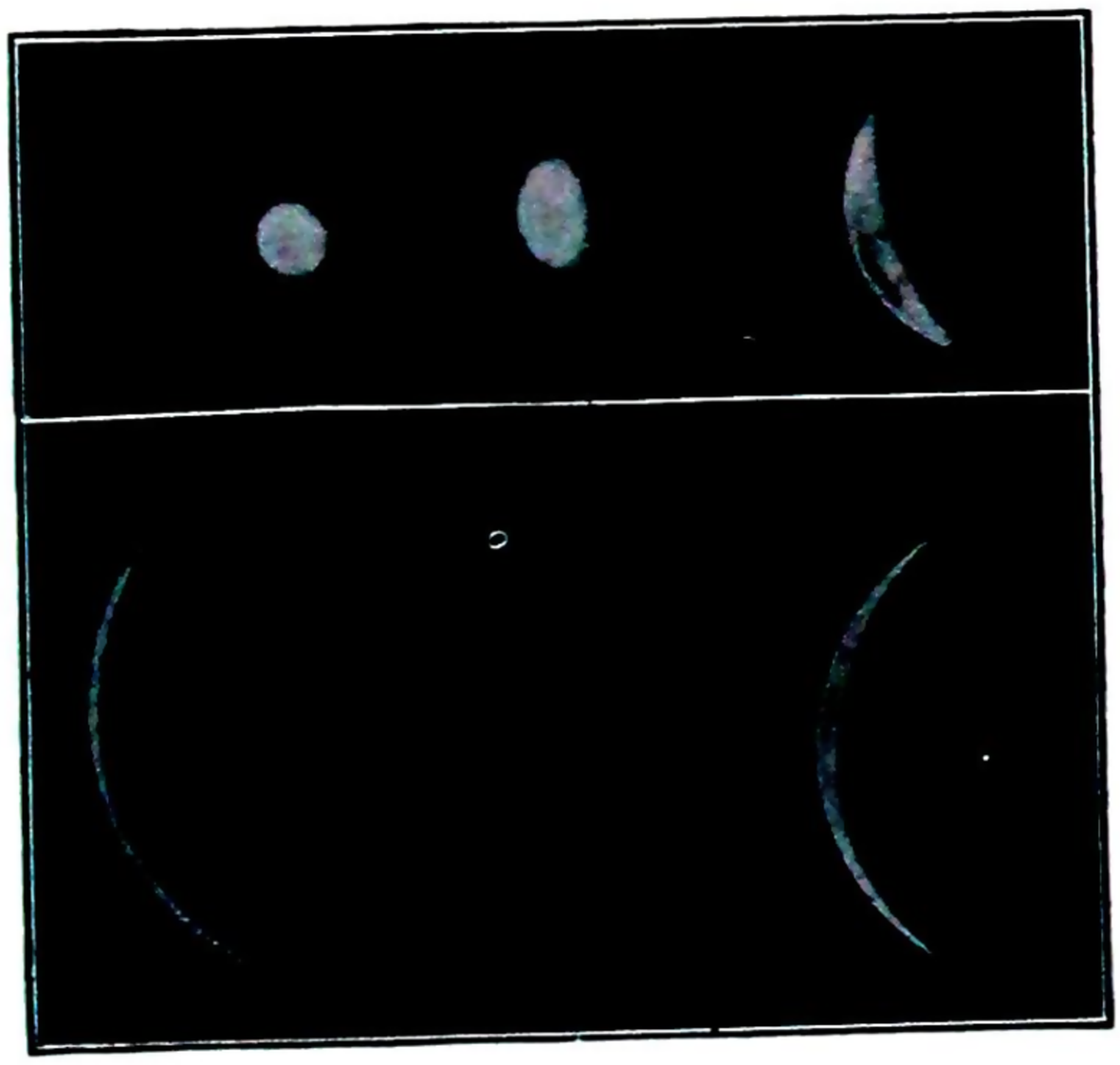


FIG. 7. Photographs of the planet Venus at different phases. (From photographs by E. C. Slipher, Lowell Observatory.)

Venus, shimmering white and surpassed in brightness only by the sun and moon, is named for the Greek goddess of beauty. The whiteness of Venus is due to the reflection of sunlight from the dense bank of clouds that always surrounds the planet, eliminating any chance of our getting a view of its surface. It will be noted from the table that Venus is much like the earth in size, mass, and distance from the sun. These facts would seem to make it a likely candidate for the honor of being the planet, other than the earth, most likely to have life upon it. On the other hand, no free oxygen has as yet been

found in its atmosphere while the carbon dioxide content of that atmosphere is many times that of the air which surrounds the earth. Thus, unless future observation should show that Venus' atmosphere does have oxygen in it, we must conclude that if any forms of life do exist on the planet they are quite different from the kinds of life that we know. There is, of course, no direct evidence for the existence of such life. As shown in Fig. 7, Venus appears to us to pass through phases similar to those exhibited by our moon. This fact, it will be recalled, was one of the first bits of experimental evidence which favored the heliocentric model of the solar system over the geocentric picture.

Earth as an astronomical body will be discussed in some detail in Chapters 6 and 7.

Mars, presenting a red appearance to the earth and named for the god of war, has received far more publicity than Venus as a possible extra-terrestrial supporter of life. There seems to be little or no direct evidence either for or against the idea, although the topic of life on Mars has, over a period of years, served nobly to fill many Sunday supplement pages. The high point in the speculation on the subject occurred in 1938 when hundreds of otherwise intelligent United States citizens misinterpreted a radio broadcast and fled their homes in terror to get away from a reported band of Martian invaders. The incident is a valuable one to remember at times when one is tempted to ridicule ancient peoples for certain of their flights of fancy. The arguments for the existence of life on Mars are based almost entirely on certain markings on the surface of the planet which some astronomers believe to be canals. The presence of such canals would, of course, imply the existence of canal builders. Other astronomers, and these are in the majority, entirely disagree with this explanation of the markings. It should be added that these marks cannot be seen with any great degree of distinctness with our present telescopes so that this is one problem on which the 200-inch instrument now under construction may throw additional light. Those who argue against the presence of living organisms on Mars base their contention on the conditions which are known to exist on the planet. For example, Mars' size is such that its force of gravity would be less than half that on earth, a fact that would affect the kind and amount of its atmosphere. Average temperatures

are much lower than on the earth, there being only a small belt which gets as warm as our temperate zones. Mars has little water and that is mostly frozen in the polar ice caps. These and other facts seem to argue almost conclusively against the existence on the planet of forms of life like those we know on earth. On the other hand, it must be recognized that they do not rule out the existence of forms of life somewhat different from those we know, forms which may have

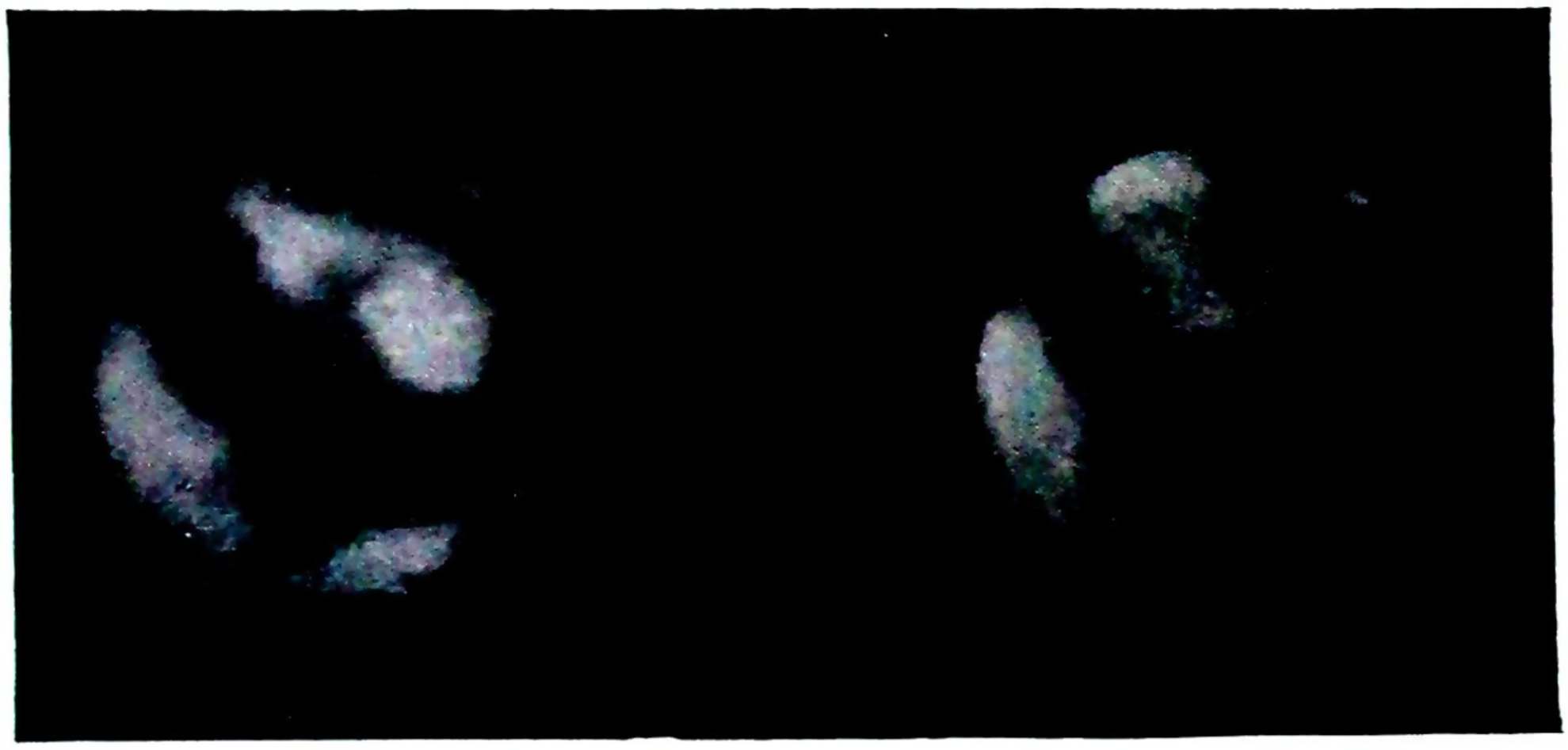


FIG. 8. Photographs of Mars taken a month apart. Note polar cap. (Mt. Wilson Observatory.)

evolved to fit the different conditions that exist on this neighboring planet.

Jupiter, the largest of the planets, is named for the king of the gods. Although Jupiter has a diameter more than ten times that of the earth, it makes one complete rotation in about ten of our hours. This rapid spinning and the fact that at least the outer layers of the planet are gaseous combine to give it a flattening at the poles much greater than that exhibited by the earth. In the matter of moonlight this planetary monarch is favored over all the others because it has 11 satellites, two having been discovered in 1938. Two of the group travel about the planet in the direction opposite to that of Jupiter's rotation and so are said to have a retrograde motion.

Saturn, the most distant and slowest moving planet known to the ancients was named for the god of time. Its day is just a little longer than that of Jupiter. In appearance Saturn varies from yellow near the equator to green in the polar regions. This planet's chief claim to

fame lies in its possession of a series of concentric rings which surround it at the equator. These rings are some 150,000 miles in width but have a thickness of less than 10 miles. The fact that stars can be seen through them testifies both to their thinness and to the further fact that they are composed of millions of bits of matter each one of which is traveling in an orbit about the planet, like a tiny satellite. In addition to the rings, Saturn possesses nine moons, the smallest of which has a retrograde motion.

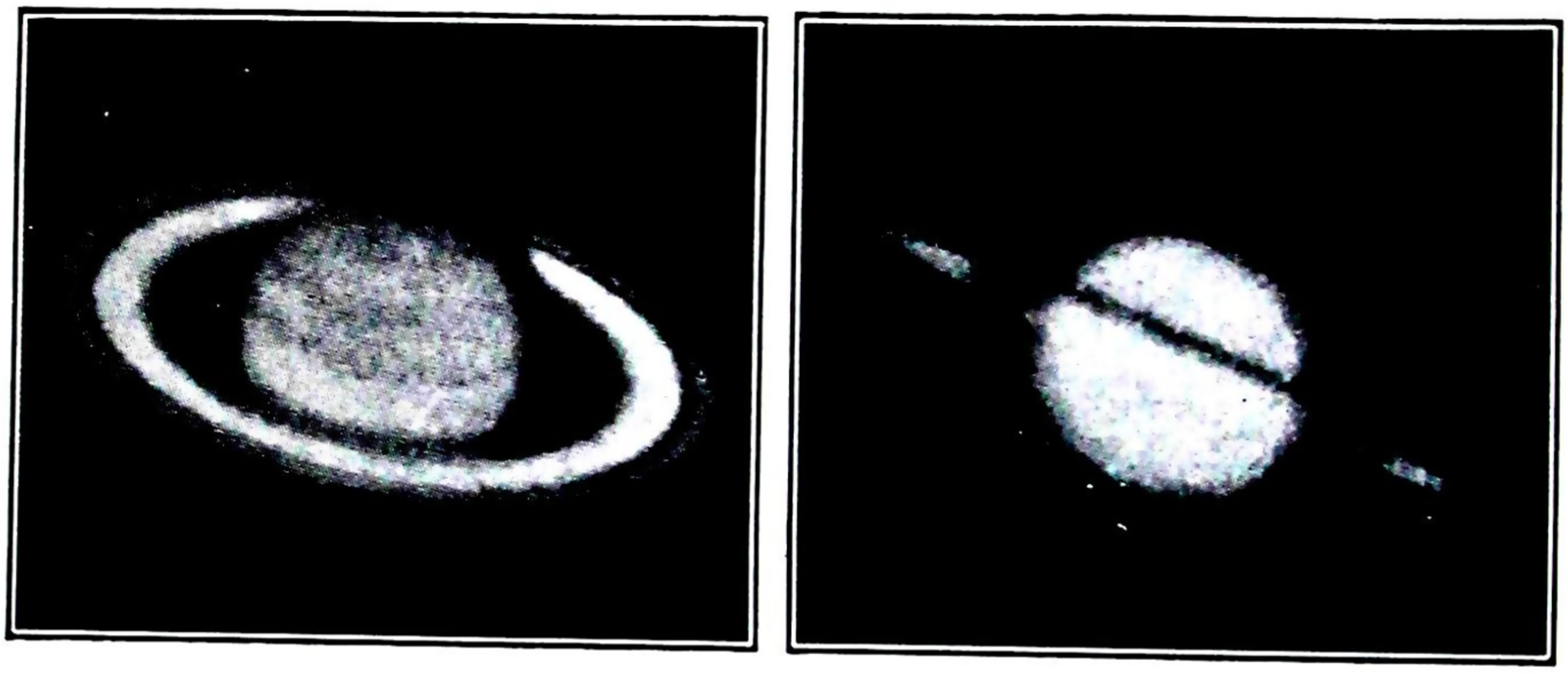


FIG. 9. Two views of Saturn with rings at different angles. Last appearance of rings in edge-wise position was in 1936. Next time maximum opening will be in 1943. (From photographs by E. C. Slipher, Lowell Observatory.)

Uranus, named for the god of the skies, is the first of the planets whose discovery occurred within modern times. The problem of distinguishing planets from stars becomes more difficult, of course, the farther away and the smaller the planet is. The five which the ancients recognized were all close enough and large enough that there could be no confusion. However, in the case of Uranus we have a body less than half as large as Saturn and over twice as far away. Thus it is not surprising that no one recognized it as a planet until comparatively recently. One night in 1781 Sir William Herschel was studying a certain stellar group. In the field of his telescope he noticed a bright object which seemed somewhat larger than a star. Continued observation indicated that it was in motion relative to the fixed stars, a circumstance which proved it to belong to the solar system. At first Herschel believed it to be a comet and so announced it, but later studies indicated its true nature. One of the most interesting things about it

is the manner in which its discovery led to the search for and discovery of Neptune and Pluto.

Neptune was discovered in 1846, after astronomers had calculated just where in space another planet should be located, and was named for the god of the sea. Immediately after the discovery of Uranus, astronomers had set to work to determine its orbit. Their results showed just how the planet should travel under the gravitational force of the sun and were corrected for the attraction of the other six known planets. By 1820, forty years after its discovery, Uranus began to show signs of departing from this prescribed path. After another 20 years had passed, the discrepancy became so great that it was said a point of light placed at Uranus' actual location and another at the place it was supposed to be could almost have been distinguished as two points of light. However, the accuracy of astronomical observation is such that astronomers were convinced of the existence of an unconsidered factor that was influencing the planet. The natural explanation was the existence of an eighth planet whose gravitational attraction for Uranus had not been taken into account. Thus, astronomers posed themselves the problem of calculating just where in all space another planet could be placed in order that it might produce the observed fluctuations in the orbit of Uranus. Such calculations, based again on Newton's law of gravitation, were made independently by Adams in England and by Leverrier in France. Their results gave locations within a half a degree of where the German astronomer Galle actually located the new planet on September 23, 1846. The success of the predictions provided the final conclusive evidence for the correctness of the law of universal gravitation. Neptune, being so very far away had to be distinguished from some 20,000 stars of equal or greater intensity.

The discovery of Pluto came in a similar fashion. In the early part of the present century Percival Lowell became convinced that Uranus' wanderings from the elliptical and narrow path calculated for her were too great to be entirely due to the gravitational blandishments of Neptune. So the search for a ninth planet began. Lowell died in 1916 after publishing his analysis of the problem, but the work was continued at the Lowell Observatory. On a photograph taken January 21, 1930, there was located the image of this distant planet,

a body so far away that it had to be separated from 15 million stars of equal or greater brightness. Because of its distance of almost four billion miles from the sun, our knowledge of Pluto is meager. There is, however, some evidence to indicate that it is extremely dense and perhaps approximately the size of Mars. Its average temperature must certainly be very low, probably in the neighborhood of 400 degrees below zero Fahrenheit, a fact which might make one wonder why this particular celestial wanderer was named for the mythological ruler of the underworld.

The Planetoids

About 1800 a number of astronomers were fixing their attention on the gap that seemed to exist between the orbits of Mars and Jupiter. Various considerations made them suspect the existence of a planet in this region. On January 1, 1801, a Sicillian astronomer named Piazzi discovered a moving body in this region while searching for a certain star listed in a new catalog, a star incidentally which later proved to be a misprint. The new discovery was immediately accepted as a planet and named Ceres for the goddess of grain. Ceres, however, proved to be only some 480 miles in diameter and later discoveries showed her to be the largest of a considerable number of small bodies that had their orbits between Mars and Jupiter. The entire group — now some 1,500 — are known as planetoids or asteroids. For a time astronomers continued to give them names, falling back on Roman deities and then Shakespearian and Wagnerian characters when the roster of Greek gods and goddesses was exhausted. Eventually matters became so complicated that the more efficient procedure of numbering the bodies was adopted.

Most of the known planetoids pursue orbits between those of Mars and Jupiter, but there are known to be several which penetrate within the orbit of Mars, while one, during part of its path, gets closer to the sun than Venus. On the other hand, there is one of the group that gets as far away from the sun as Saturn during part of its year. In diameter they range from a few miles up to the 480-mile thickness of Ceres. It has been suggested that they may represent the fragments of what once composed one or more larger bodies — bodies which met with some early astronomical catastrophe. Such a hypothesis is not

unreasonable in the light of our present knowledge, although it has no actual experimental evidence supporting it.

Comets

Of the hundreds of comets so far observed and recorded, only a few have been visible to the naked eye. However, the spectacular and often terrifying appearance of these few made them objects of great interest and speculation to early man. He believed them to be sent by a supernatural power to punish him or warn him of approaching disaster. It was only after centuries of observation and collection of evidence, that he came to realize that the appearance of these objects represented a perfectly natural phenomenon.

A typical comet, as we observe it, consists of a nebulous head containing a bright nucleus of starlike appearance and one or more streamer-like tails extending out sometimes as far as 100,000,000 miles. Comets whose paths have been charted travel in orbits which are ellipses, although much more flattened than those of the planets. Since comets obey Kepler's second law, it is obvious that only a small fraction of their time is spent anywhere near the sun.

The part of the comet which we call the tail seems to be composed of nebulous material which has been forced out of the head by the pressure of radiation from the sun. Part of the evidence for this lies in the fact that the tail is always found on the opposite side of the head from the sun and only becomes visible as the comet approaches this stellar member of the solar system. Thus as a comet recedes the tail goes before the head instead of following it.

The best known of all comets is the one called Halley's comet. The relation of its orbit to those of the planets is indicated in Fig. 4 (page 24). Edmund Halley, an English astronomer and friend of Sir Isaac Newton, was the first man to suspect that this comet pursued a regular path returning periodically within the range of man's vision. He calculated its period and predicted correctly its appearance in 1690. After having determined the period as being approximately 75 years, Halley found records of previous appearances of the same comet dating back to 240 B.C. In 1066 it had been heralded as being connected with the victories of William the Conqueror. Its 1456 showing occurred shortly after the fall of Constantinople and the prayer of the

Christian world at that time was, "Save us from the Devil, the Turk and the Comet." Its most recent engagement was played in 1910 at which time it came so close to us that its tail swept over the earth, an occurrence which seemed to produce no observable effects.

The comet recently discovered by Leland S. Cunningham of the Harvard College Observatory is shown in Fig. 10. Its appearance marks the first time since 1910 that a comet has been conspicuously visible to the naked eye. In that year there were two such, one being Halley's Comet. The Comet Cunningham reached its closest point to the earth around January 10, 1941, at which time it was some 54,000,000 miles distant from our planet and about 33,000,000 miles from the sun.

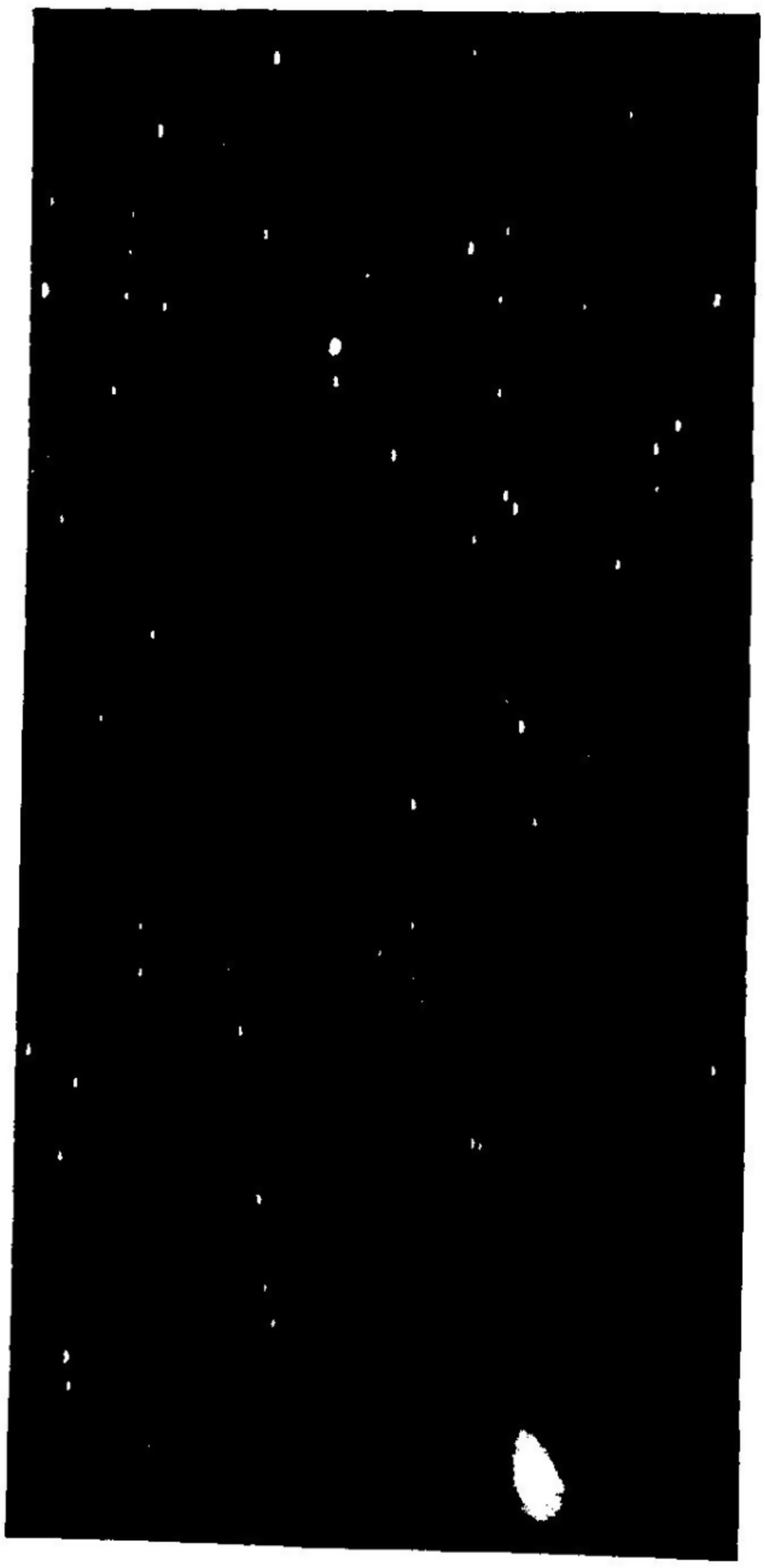


FIG. 10. Comet Cunningham photographed with 24-inch reflector, January 2, 1941. (Yerkes Observatory and University of Chicago Press.)

If, as has been suggested, the tail of a comet is made up of material which has been forced out of the head, each trip around the sun should find the head somewhat smaller and eventually the comet might be expected to disappear altogether. Actually a few cases are known where a comet was charted for several successive appearances and then never seen again. Another bit of evidence which supports the above hypothesis of the nature of the cometary tail will be mentioned in connection with our consideration of meteors and meteorites.

Meteors

The last group of bodies which must be included as members of the solar family are the meteors, commonly but incorrectly called

shooting or falling stars. Everyone is familiar with the appearance of these objects, some of which surpass Venus in brightness during their brief incandescent career. Millions of meteors are believed to enter the earth's atmosphere every 24 hours, most of them so small that several thousand could be held in one's hand. As they enter the earth's sphere of gravitational influence they are drawn toward it. At heights



FIG. 11. Great Meteorite, Grootfontein, South Africa. The largest known meteorite, weighing more than 50 tons. Exposed surface is about 9 x 8 feet. (Courtesy W. J. Luyten, Harvard College Observatory.)

of from 60 to 80 miles above its surface these objects encounter atmosphere of such density that friction raises them to incandescence. Then they become visible to us and remain so until they either are burned up, as are the vast majority, or strike the earth.

The few that reach the surface of the earth before being consumed are called meteorites. Most of these are masses of stone containing traces of iron while a small percentage are composed of almost pure iron and nickel alloyed together. Several hundred meteorites are believed to strike the earth every year, of which three or four on the average are recovered. The largest one on display is the 36-ton specimen brought back from Greenland by Admiral Peary and now located in the American Museum of Natural History in New York. Figure 11

shows one of the largest meteorites that has been found. In 1908 a gigantic meteorite fell in Siberia, completely devastating an area of over 1,000 square miles. Meteor Crater in Arizona marks the place at which a huge meteor is believed to have landed in prehistoric times. Either the main body of the meteorite is buried so far beneath the surface that it has not been reached or else it exploded into small fragments at the time it fell. However, discovery of large amounts of meteoric material around the crater leaves little doubt as to what caused the crater.

Although many of these bodies are believed to enter our atmosphere sporadically from outer space, the fact that great showers of meteors become visible at definite intervals of time indicates rather conclusively that large numbers of these fragments pursue definite orbits within the solar system. Also it has been shown that the orbits of many of the showers coincide with the paths of well-known comets, a circumstance which suggests that the meteors in the showers originated as material forced out of the head of a comet and coming within our atmosphere at intervals when the earth's orbit intersects the orbit of the comet.

The Origin of the Solar System

As long as man's actual knowledge of the solar system was small, so was his urge to speculate and theorize concerning its origin. In general he was perfectly satisfied with the simple statement that a Supreme Being made it as it is. But as time went on and he came to know more and more about the manner in which these heavenly bodies performed, every discovery added to the evidence that life in this solar family goes forward in an exceedingly orderly fashion. The planets traverse fixed orbits in definite and unvarying intervals of time; day and night, the moon's phases, the seasons, solar eclipses, are all definite, periodic, predictable phenomena. In fact as far as man could discover, all of these celestial events were absolutely dependable and none of them came about sporadically or haphazardly. Consequently, he began to reason that it would be only natural to expect that there had been an orderly, cause-and-effect series of events leading up to the present organization of the solar system. And so he began to search for a hypothesis or explanation which would account in a reasonable

manner for the present arrangement of things. Again it should be emphasized that one's adoption of a belief in an orderly origin of the solar system in no way necessitates abandoning one's belief in the existence of a God.

The Nature of the Problem

Before we examine two of the hypotheses that have been advanced as possible ways by which the solar system as such came into being, let us review briefly some of the experimental evidence with which any such theory must agree. Recall also that one lone experimental fact that is definitely contradictory to a scientific theory may be sufficient to overthrow it completely, although ten million facts that support it cannot be said to prove it absolutely, since we have no way of knowing what new facts may be discovered next week or next year. Indeed, it may be that to inquire whether or not a hypothesis is correct is to ask an essentially meaningless question. Dr. W. F. G. Swann has said that such a question makes about as much sense as to ask if the Chinese language is correct. He says that in both cases one should ask rather whether they are satisfactory. The Chinese language is satisfactory if it provides a tool with which communication can be carried on among Chinese people and a scientific theory is satisfactory if it provides a language in which known experimental phenomena can be described. Thus it may be absolutely satisfactory today but absolutely unsatisfactory a year from today, if contradictory evidence is discovered in the meantime.

The data with which any picture of the origin of the solar system must agree include the facts that there are at least nine planets and some 1,500 planetoids which, with no exceptions, travel about the sun in the same direction; that as far as we know the sun and all the planets except Uranus rotate in the same direction; that of the 28 known satellites 20 revolve in the direction of rotation of the parent planets while eight have the retrograde motion previously defined; that Kepler's laws describe accurately the motions of the planets; that there is a force of attraction between objects which is described correctly by Newton's Law of Universal Gravitation.

Then in the light of these and additional experimental facts, one asks what sort of a series of events might have occurred that would

bring about this present organization. Of the many hypotheses that have been advanced from time to time, we will examine two, one which formerly was widely accepted but which was overthrown by subsequent evidence, and one which is accepted today as being reasonably satisfactory in the light of present knowledge.

The Nebular Hypothesis

In 1796 the French astronomer and mathematician Pierre La Place advanced a hypothesis concerning the origin of the solar system, a hypothesis which is known variously by the names nebular, ring, or La Placian. The theory La Place himself said he put forward as " conjecture which I present with the distrust which everything not the result of observation or of calculation ought to inspire."

The nebular hypothesis traced the events leading up to the formation of the solar system as follows:

" At one time all the matter now in the solar system was in one huge nebulous mass or cosmic cloud which occupied more space than that now included within the outermost planetary orbit. This cloud was at a high temperature and was rotating. As it cooled and contracted strains were set up within it until finally a break occurred. This break resulted in splitting off a ring of material which later coalesced to form the outermost planet. The tension thus relieved, the cloud continued to rotate and contract until new strains were set up and the splitting off procedure was repeated. This continued until all the planets were formed, the central mass remaining being the sun. The satellites of the planets were formed in the same fashion, as the material of the planets themselves cooled and contracted."

The principal bits of evidence advanced for this hypothesis were two. First the fact of the gaseous nature of the sun agreed with the picture although it could scarcely be said to prove the theory as a whole. The second and more striking evidence offered concerned the appearance of the planet Saturn. If one accepted the theory, then the rings of Saturn could be accounted for beautifully as satellites in the process of formation. This was not particularly sound evidence since in order to use it as such, one had first to assume the correctness of the theory which one was trying to prove. Nevertheless, it was impressive and as long as there was little or no contradictory evidence, the hypothesis was widely accepted.

Contradictory evidence soon began to accumulate to justify the distrust originally expressed by La Place himself. One difficulty that had existed all along was the lack of any earthly precedent for rings breaking away in order to relieve tension in a rotating object. If the strain becomes too great in rapidly rotating bodies such as fly-wheels, pieces always fly out. Also there seemed to be no mechanical explanation for the building of such instabilities in the first place. Then the discovery of satellites with a retrograde motion about the planet introduced additional difficulties. There seemed to be no good reason why the rings, after they split off, should stop and then begin to revolve in the opposite direction. Also satellites were discovered which completed a trip around the parent planet in less time than it took the planet to rotate once, and it seemed unlikely that the ring would speed up after splitting off. Eventually the mass of contradictory evidence became so great that the hypothesis was abandoned.

The Planetesimal Hypothesis

In connection with the nebular hypothesis we saw that it was impossible to explain the occurrence of the periodic instabilities in terms of forces within the rotating cloud itself. But such strains might be set up as a result of the action of an outside agent of some sort. The postulation of such an agent forms the basis of the planetesimal hypothesis advanced in 1900 by Dr. T. C. Chamberlain and Dr. F. R. Moulton, geologist and astronomer respectively, at the University of Chicago.

The planetesimal picture tells a somewhat different story of the embryonic development of our system of sun and planets. To begin with, essentially all the material now in the system is again assumed to have once been one huge star body. This star is pictured as having been visited by another one some billions of years ago. The visiting star did not collide with ours but simply passed by, as it were, within astronomical waving distance. The force of its gravitational attraction, however, set up mammoth tides on the sun's surface. The star masses and assumed separations were such as to make it reasonable for huge chunks of matter to have been flung out into space from the sun as a result of these tides. Along with the larger bodies which are thought of as becoming the planets, undoubtedly hundreds of smaller

fragments would have been turned loose in space as the breaks occurred. Some of these drifting within the gravitational influence of planets could reasonably be expected to become satellites and follow orbits about them. Since their directions of revolution would depend solely on how they happened to be moving when caught by planetary attraction, the existence of moons with a retrograde motion offers no difficulty. The planetoids and planets, however, should all have the same direction of revolution about the sun, as they do. Altogether it may be said that at the present time there seems to be no evidence which absolutely contradicts such a hypothesis, although it has a few weak points. A number of variations on the above basic picture have been advanced since 1900 but need not be discussed here. All follow a fundamental outline like the one just described. Consequently, a hypothesis of this sort is generally accepted as offering a reasonably satisfactory picture of the events leading up to the formation of the solar system, in the light of our present knowledge.

Questions

Discussion

1. Name all the different types of celestial bodies that are represented in the solar family.

2. What are the reasons for thinking that there may be life on Venus and what actual evidence is there of the existence of such life?

3. List all the arguments you can find both for and against the existence of life on Mars.

4. See if you can work out an explanation for the fact that Venus and Mercury are the only planets which exhibit phases.

5. How does the time from noon to midnight on the earth compare with what it would be for an inhabitant of Neptune? Of Jupiter? Of Mars? Of Mercury?

6. Explain how an astronomer distinguishes planets from stars when he studies the heavens.

7. At which of the following times would you expect Mercury to be most easily observed and why? At noon. At midnight. At dawn.

8. What is the basis for classifying the planets into two groups?

9. Describe the discovery of Neptune and Pluto.

10. In what respects are planetoids like planets and in what ones do they differ?

11. Give some evidence to support the theory that the tail of a comet is composed of material forced out of the head by solar radiation pressure.

12. Look up information about one or two specific comets other than those discussed in this chapter.

13. Discuss the connection that is believed to exist between comets and showers of meteors.

14. State the basic assumptions of the nebular hypothesis.

15. State the basic assumption of the planetesimal hypothesis.

16. Explain how the following observed facts served as evidence for or against the nebular hypothesis, the planetesimal hypothesis, or both:

a. The rings of Saturn.
b. Retrograde satellites.
c. The gaseous nature of the sun.
d. A satellite whose period of revolution about the parent planet is less than the latter's period of rotation.

Multiple Choice

1. List all of the following that may be said to be members of the solar family. (a) Venus. (b) Halley's Comet. (c) Alpha Centauri. (d) The Big Dipper. (e) Ceres.

2. Which one of the following facts seemed impossible of explanation on the nebular hypothesis? (a) The star closest to the solar system is over four light years away. (b) All planets revolve about the sun in the same direction. (c) The outer moon of Jupiter revolves in a direction opposite to that of rotation of the parent planet. (d) Saturn is surrounded by a series of rings which are very thin in comparison to their width. (e) The path of Uranus deviates slightly from a true ellipse.

3. The tail of a comet (a) forms when the comet is farthest from the sun and gradually disappears as it moves nearer, (b) seldom has a maximum length of more than a few thousand miles, (c) is so named because it always follows the head, (d) always precedes the head, (e) reaches its maximum length when the comet is nearest the sun.

4. Both comets and meteors (a) are composed of solid matter rather than gaseous, (b) must enter the earth's atmosphere before we can see them, (c) are raised to incandescence by friction with the earth's atmosphere, (d) return periodically, (e) shine principally by reflected sunlight.

5. Of the members of the solar system, an observer located on Alpha Centauri, if he were about like ourselves, would be able to see (a) the sun and major planets, (b) none whatever, (c) the sun only, (d) the sun and Jupiter only, (e) the sun and all the planets.

True–False

1. One who accepted the nebular hypothesis would probably believe that Mercury was the first planet to be formed.

2. The rings of Saturn do not now constitute a stumbling block to the planetesimal hypothesis.

3. Planetoids resemble planets in that they are not luminous themselves.

4. It is generally accepted that the passage of the earth through the tail of Halley's comet in 1910 caused a series of disastrous hurricanes in various parts of the earth.

5. Every element ever found in meteorites has also been found on the earth.

6. The closer a comet gets to the sun, the more rapidly it travels.

7. The brilliant white appearance of Venus leads us to believe that its surface is bathed in perpetual sunlight.

8. If Neptune were suddenly to be plucked out of the solar system, the orbit of the earth would change to a greater or less extent.

9. It is correct to say that present-day evidence proves conclusively that the planetesimal hypothesis is a true picture of the formation of the solar system.

10. The planet whose physical characteristics most nearly resemble those of the earth is Mars.

Chapter 4

THE UNIVERSE BEYOND THE SOLAR SYSTEM

THUS far our discussion of the universe has been confined to what may be called our own astronomical backyard. In this tiny fragment of space we have located one star, nine planets and their satellites, some 1,500 planetoids, and a number of comets and meteors. Now we are about to become somewhat more cosmopolitan in our cosmic outlook and concern ourselves with what lies beyond the solar system. For after all, the orbit of the outermost planet has a diameter of a mere seven and a half billion miles which is just a little more than a thousandth of a light year, whereas many of the stars one can see with the naked eye are hundreds and even thousands of light years distant.

OUR GALAXY

Its Nature and Size

The work of various astronomers performed during the latter part of the 19th century seemed to indicate that the sun is but one of a vast cloud of stars, a cloud which is called a galaxy, or galactic system. In the light of the earlier of the evidence, astronomers pictured this cloud as having approximately the shape of a disk, with fuzzy irregular edges. It was believed that from a point without, say, a few hundred thousand light years away, it would have much the appearance of a huge, light-colored, disk-shaped, swarm of bees. At the present time astronomers are inclined to think that, observed directly from the side, the galaxy would look more like a spiral than a simple disk. Such a modification, however, does not materially affect the dimensions given below. This star swarm or galaxy is believed to have a diameter of from 150,000 to 200,000 light years and a thickness of some 20,000 light years. Our solar system is located at about the

center along the short dimension, and approximately one-fourth of the distance from the axis of the disk or spiral toward the circumference. To give a comparative picture of the relationships involved we may think of the distance across the galaxy as being set equal to the diameter of the earth. On such a scale its thickness would be about the equivalent of the distance from New York to Chicago. The distance across Pluto's orbit, which essentially defines the limits of the solar system would, on our model, be some 2½ inches, while the diameter of the sun would be less than the thickness of a sheet of paper. From the sun to Alpha Centauri, its nearest star neighbor, would be a distance about equal to the height of an eight- or nine-story building.

We may seem to have been rather highly speculative in building up this picture of how the galaxy would appear to someone outside it. Let us see if we can to some degree substantiate our hypothesis by asking how such a congregation of stars should appear to observers like ourselves who are inside the system.

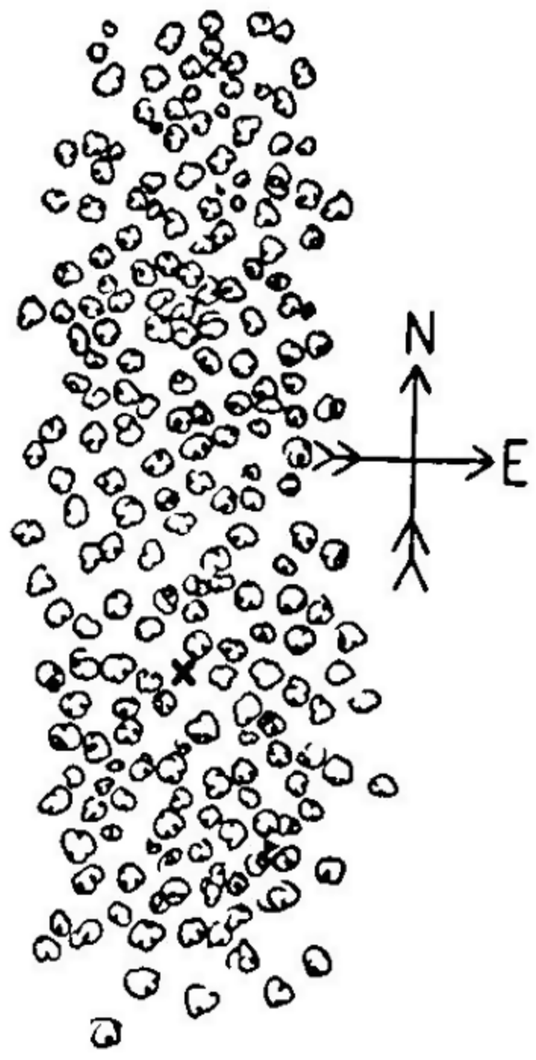


FIG. 12. Diagram showing a man located in a long narrow forest at a point somewhat analogous to the position of the solar system in our galaxy.

By way of illustration we will direct our attention first to Fig. 12 which represents a long narrow forest within which a man is located at the spot marked, in the orthodox manner, by *X*. Concerning this man, we ask what sort of an idea of the forest he might be expected to get if he looked about him in various directions. If he looked either directly north or directly south, the trees would appear very dense, so dense that if he was a little near-sighted and unable to determine that the trees were varying distances from him, he might very easily think of them as being so close together as to form almost a solid wall of wood. On the other hand, when he turned his gaze either to the east or west, he would see a considerable amount of daylight since the distance to the edge of the forest is so much less. Thus he might conclude that the forest was much less dense in these directions than to the north or south, although the actual number of trees per acre was about the same throughout the tract.

Now our position in the galaxy is believed to be very much like that of our near-sighted friend in the forest. If so, an earthly observer, gazing out through space in the direction of the long dimension of the galactic system, would have vast numbers of stars between him and the boundary of this stellar cloud. In view of the great number, they might be expected to appear to overlap and run together giving

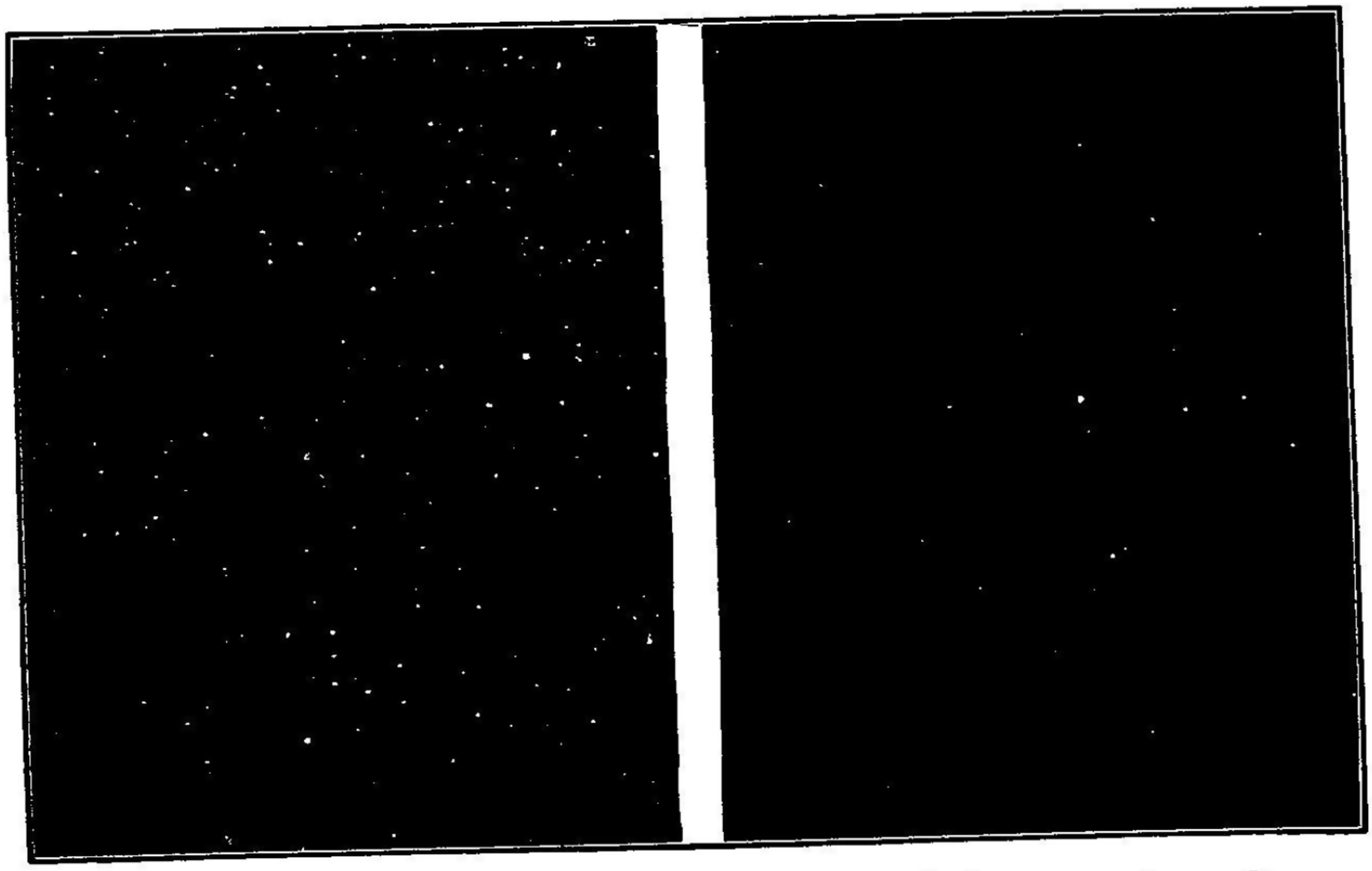


FIG. 13. Photographs of two regions of the sky of the same size. One was taken toward Milky Way and other at right angles to that direction. Faintest stars in both regions of 18th magnitude. (Mt. Wilson Observatory.)

the effect of a cloudlike streak extending across the sky. The view along the short dimension, however, should show no such phenomenon since the edge of the galaxy is so much closer to the observer. Now note the photographs shown in Fig. 13. It will be seen that the one taken toward the Milky Way appears exactly as was suggested above for a view in the direction of the long dimension of the galaxy, while the other is just the sort of picture one would expect to see in the direction of the short dimension. Thus the appearance of the heavens seems to confirm this hypothesis concerning the shape of the galaxy, particularly since modern telescopes demonstrate clearly that the Milky Way is composed of individual stars.

Members of the Galaxy

Observation indicates that there are two kinds of members of the galaxy. In addition to the stars mentioned above there are somewhat hazy appearing bodies called nebulae. We will consider each of these types briefly.

Stars. — It is difficult to estimate accurately the size of the star population of our galaxy, because of the fact that whenever a larger telescope has been built, the number of stars that are visible has increased. It is believed that with the largest one now in use, the 100-inch instrument on Mt. Wilson, some two or three billion can be photographed, just about enough, in other words, so that every man, woman, and child in the world could have one named for him. There are certain considerations that lead astronomers to think that this represents about one-fiftieth of the total number in the galaxy. Of these billions of stars, only some 5,000 or 6,000 are close enough to be seen without the aid of a telescope and, of course, not over half of these at any one time, since the earth itself blots out one's view of the rest.

In spite of the enormous number of stars in the galaxy and the congested appearance of the Milky Way, actually there is no overcrowding. The volume of the galactic system is so great that the sun and all its millions of brothers have plenty of room in which to move about. Sir James Jeans has said that the density of the star population in the galaxy could be fairly compared to wasp " crowdedness " in a hollow cube, 1,000 miles on a side, in which six of these insects were flying about at speeds equal to a thousandth of a snail's pace.

Even a hurried glance at the sky on a clear night shows that there is great variation in brightness among the stars. The ancient astronomer Hipparchus recognized this and set up a system of classification of these bodies into magnitudes. The twenty brightest stars he put in the first magnitude and the faintest ones he could see, in the sixth. The system was continued to more magnitudes as the telescope was invented and improved, until at the present time the faintest stars listed are of the twenty-first magnitude. These magnitudes bear a rather peculiar relationship to each other. As the measurement of brightness became more accurate it was found that Hipparchus' first magnitude stars were about 100 times as bright as those in magnitude six. In

other words, a spread of five magnitudes meant a change in brightness of one hundredfold. A spread of one magnitude, therefore, would have to be represented by the number which multiplied by itself five times equals 100, that is the fifth root of 100. This number happens to be the rather inconvenient one of 2.512 which means that a star of any magnitude is 2.512 times as bright as a star of the next magnitude.

TABLE 2.[1] THE FIFTEEN BRIGHTEST STARS VISIBLE AT LATITUDE 40° N

Name of Star	Constellation	Magnitude	Distance in Light Years	Luminosity Compared with Sun	On Meridian 8:30 P.M.	
					Date	Degrees from Zenith
Sirius	Canis Major	—1.58	8.8	26.3	Feb. 20	S 57
Vega	Lyra	.14	26	50	Aug. 19	S 2
Capella	Auriga	.21	52	165	Jan. 25	N 5
Arcturus	Boötes	.24	41	100	June 13	S 21
Rigel	Orion	.34	500	16,000	Jan. 25	S 49
Procyon	Canis Minor	.48	10.5	5.5	Mar. 3	S 35
Altair	Aquila	.89	16	9.2	Sept. 8	S 32
Betelgeuse	Orion	.92	200	1,200	Feb. 4	S 33
Aldebaran	Taurus	1.06	57	90	Jan. 15	S 24
Pollux	Gemini	1.21	32	28	Mar. 4	S 12
Spica	Virgo	1.21	230	1,500	June 1	S 51
Antares	Scorpio	1.22	380	3,400	July 16	S 67
Fomalhaut	Piscis Australis	1.29	24	13.5	Oct. 28	S 70
Deneb	Cygnus	1.33	600?	10,000?	Sept. 23	N 5
Regulus	Leo	1.34	56	70	Apr. 12	S 28

[1] Richards, *The Universe Surveyed*, p. 610, Table 2. New York: D. Van Nostrand Co. 1937.

Still better techniques demonstrated that these twenty brightest stars varied considerably among themselves in brilliancy. Now if a certain value is taken as standard first magnitude brightness, then a star that happens to be 2.512 times as bright as this must be put in zero magnitude and one still brighter can only be represented by a negative magnitude. And so we have the rather odd fact that Sirius, the Dog Star and brightest star in the heavens, is of magnitude —1.58. It should be noted that these magnitudes represent only apparent brightness, i.e., how bright the body looks to us. Since this is affected by distance as well as actual brightness, magnitude is not representative of the true brilliancy of any star. This will be evident from a comparison of the magnitude, distance, and luminosity columns in the table.

The last two columns in Table 2 give data for the location of

these fifteen stars. Under the one labelled " Date " is listed the night of the year when each one is on the meridian at about 8:30 P.M. The other column then indicates how far north or south of the zenith the star is located. Instructions for locating at any time of the year some of these fifteen as well as certain others will be mentioned a little later in the chapter.

A considerable number of stars are constantly changing in brightness and are known as variables. For the most part these variations are periodic, although the lengths of the periods differ widely, some fluctuating rapidly over a small range of brightness and others requiring months to complete a cycle. The star Mira, for example, varies from a magnitude many times too faint to be seen with the naked eye to a brilliancy about equal to that of the Pole Star and back again, taking over 10 months for the cycle. Perhaps the most interesting group of stars which vary in brightness are the so-called Novae or " new stars." Actually they are not young at all but respectable adult members of the star community who for some reason or other suddenly go on a grand spree of brilliancy. Within a few days or weeks such a star may increase many fold in brightness only to fade away again to its old magnitude. A fourteenth magnitude star, known as Nova Herculis, in 1934 suddenly began to grow brighter and within two weeks had become a star of magnitude two. A year later its brief day in the sun was over and it had returned to magnitude fourteen. One was recorded in 1572 which at its brightest was visible in the daytime. An average of two or three novae are discovered each year. The reason for the occurrence of these sudden and strange stellar outbursts is as yet unknown.

Stars are of a variety of sizes and colors, the latter characteristic being an indication of surface temperature. Thus the temperature of dull red stars is believed to be about 2,500° F., of orange ones some 5,000° F., of yellow stars approximately 10,000° F., while the bluish-white members of the stellar family are thought to be as high as 35,000° F. and higher. These two properties of size and color were linked together in an interesting theory about the life history of stars, advanced about 1913. This hypothesis pictured an embryo star as being a large, diffuse mass of gas of low density and at a low temperature. Throughout its life, according to the theory, the star decreased in size, finally becoming what was called a " dwarf " although it had begun life as a " giant." The temperature changes occurring in it were be-

lieved to be such that it increased in temperature during the first half of its life, changing in color from dull red through orange and yellow to bluish-white at middle age. This color and temperature sequence was believed to be reversed during the last half of the star's life cycle so that when its career came to an end it was dull red again. Thus the so-called " red giants " were pictured as being very young stars and the " red dwarfs " as very old stars.

Experimental evidence acquired since 1913 has made it clear to astronomers that this particular picture is no longer a satisfactory one. Although stars are still believed to pass through a life cycle of some sort, theories concerning it have been in a state of change for some years, and at the present time there is none that is completely satisfactory and generally accepted.

Stars are all in rapid motion through space, although we frequently speak of them as being fixed. An airplane traveling 150 miles per hour appears to be traveling very rapidly if it is only 100 feet above the ground whereas the same plane moving at the same speed at an altitude of 10,000 feet may seem almost motionless. In like fashion the enormous distances which separate us from the stars cause them to appear to be stationary with respect to each other. Over long periods of time, of course, their relative position should change and we know from ancient drawings that various constellations do have a different appearance to us than they did to astronomers of earlier times. One might speculate on the problem that exponents of the pseudo-science of astrology will have a few tens of thousands of years from now when the natural motions of the stars will have jumbled up and made unrecognizable the constellations by which they set such store.

In addition to the foregoing, there is one other type of motion which about a fifth of the stars possess. These particular stars are in pairs and revolve about each other. In view of the fact that there are so many such pairs, known as double stars, it has been suggested that the phenomenon may simply represent a stage in the evolution of some stars. It may be that the speed of rotation of some single stars becomes large enough to cause a piece to be flung out and take up a path about the parent body. The fact that in most cases the stars are of unequal sizes and frequently are at different temperatures would seem to support this idea.

Nebulae. — Also included within our galaxy are a number of bright, cloudlike masses known as nebulae. These are given names such as planetary, elliptical, diffuse, and the like, depending on the way in which they appear to us. A diffuse nebula is shown in Fig. 14. It was thought at one time that better telescopes would reveal these to



FIG. 14. A diffuse nebula in the Pleiades. (Mt. Wilson Observatory.)

be made up of individual stars, as is the Milky Way. Such has not been the case, however, so that now they are believed to be simply masses of dust and gas, some shining by reflected star light and others by their own luminescence. One suggestion, not proven as yet, is that they may represent early stages in the formation of stars.

Another kind, called dark nebula, are phenomena that no one has ever seen but which have been assumed to exist in order to account for certain other observed facts. There are a number of large black

patches in the sky, areas in which no stars at all can be seen (Fig. 15). This means either that there are no stars in these regions or that something is between us and the stars, blocking off our view. Since there seems to be no good reason why such large sections of space should be entirely devoid of these bodies the second hypothesis seems the more

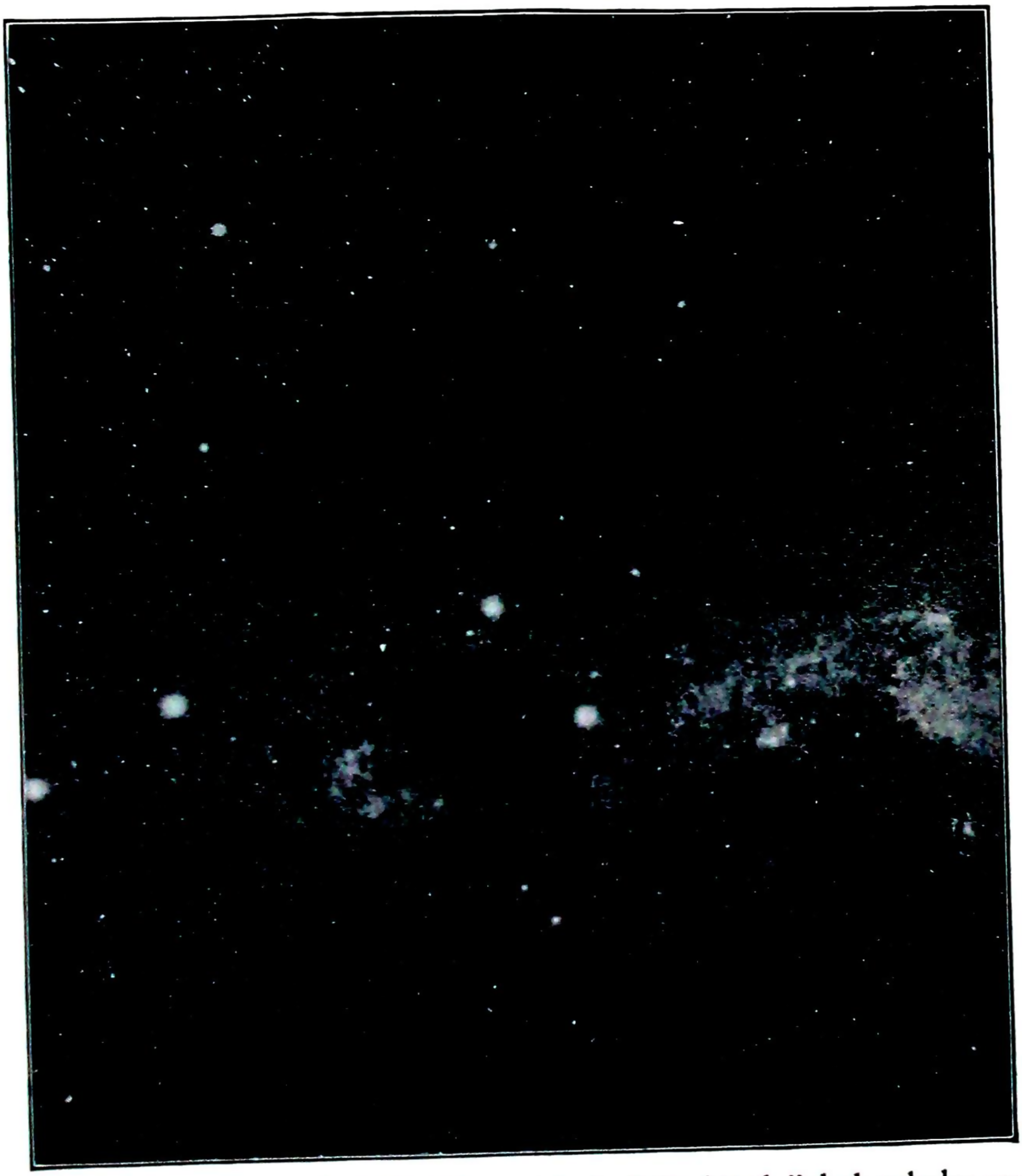


FIG. 15. A section of the Milky Way with the " Coal Sack " dark nebulae near the center. Near it is the Southern Cross. Bright star at extreme left is Alpha Centauri. (From a photograph by Margaret Harwood at Arequipa station of the Harvard Observatory.)

reasonable of the two, and nebulae which are dark and opaque are assumed to screen these parts of the heavens. Between 300 and 400 of these dark nebulosities have been observed.

Stars and Constellations

To end our story of the stars now, after discussing the dry data concerning their number, brightness, age, size, and temperature would be to ignore completely one of the most interesting phases of the picture. Much of the fascination that the stars hold for hundreds of people lies in the game of locating the constellations in the heavens. While space does not permit the charting of an exhaustive stellar journey, it may be that a map and a few directions on how to start will arouse some desire in the reader for further astronomical touring.

It was many centuries ago that man first noticed the constant relationship that the stars seemed to maintain with respect to each other. Although they might rise, move across the sky, and set each night, or disappear entirely during portions of the year — effects which we know are caused by the earth's rotation and revolution respectively — yet whenever he could see them they always formed the same pattern. Thus it was natural that he should give names to the groups, link them with his earthly affairs, and weave legends about them.

Of the 88 star groups or constellations which now are listed, 48 were chosen by early man. These abound in animal names and include two bears, several dogs, a lion, a bull, a ram, several fish, a dragon, an eagle, a swan and others. Back of most of these titles are legends and stories which taken together make the heavens a sort of mythological history book. The remaining 40 constellations have been given their names in modern times.

Directions for Using the Star Charts

Inside the covers of the book there are located star maps of the northern and southern hemispheres of the sky. The outer edge of the circle in each case represents what may be called the sky's equator, that is, the great circle of the heavens directly above the earth's equator. The centers of the two charts are the points at which the earth's poles point.

To use the northern sky map, the reader should proceed as follows:

1. Rotate the chart until the date at the top of it is the one nearest that on which the map is to be used.

2. To establish the horizon, place a straight-edge across the map horizontally and below the center at a point such that it is tangent to the circle which corresponds to the latitude of the observer's location.

3. The chart is now properly oriented for about 8 o'clock in the evening. For each hour after 8 o'clock the map should be rotated in a counter-clockwise direction through one of the half-month divisions.

The manipulation of the southern sky chart differs from the above in two particulars; namely,

1. The horizon line is placed above the center tangent to the proper latitude circle, instead of below, and

2. For each hour after 8 o'clock, the map is rotated clockwise instead of counter-clockwise.

The Great Bear

Since most people know where in the sky to look for the Big Dipper, we shall make it the starting point of a short sky tour. This group of stars is part of a constellation known as Ursa Major or the Great Bear. The seven stars which constitute the Big Dipper are the brightest in the constellation and are, on the average, some 70 light years distant.

One striking thing about this particular constellation is that it was given the name bear by many ancient peoples, so widely separated geographically that it is inconceivable that there could ever have been any communication between them. Olcott [1] says,

" On the banks of the Euphrates thousands of years ago it was so designated and the Iroquois Indians of North America called this star group " Okouari " their name for " bear." The Algonquin Indians called the constellation " the Bear and Hunters " . . . The Finns called Ursa Major " Otawa," a title resembling the " Okouari " of the Iroquois . . . Thus in remote parts of the earth, in the far north, from the valley of the Euphrates to the region of the Great Lakes of North America we find the same stars likened to an identical animal."

1 Olcott, W. T., *Star Lore of All Ages*, p. 3. New York: G. P. Putnam's Sons, 1911.

There seems to be no completely satisfactory explanation for this phenomenon. Aristotle suggested that it might have come about because the bear was the only animal known to the ancients that could live as far north as the constellation seems to be. However, it is difficult to see how some of these peoples, such as the inhabitants of the Euphrates valley, would know that bears could live in the far north. The pattern of stars in the group does not seem to suggest a bear any more than almost any other animal so that cannot account for it. As a matter of fact, there is no reason to think that the constellations were always named because the stars seemed to form the outline of a certain animal. Unquestionably they were often named simply in honor of a particular bird or animal or legendary character and so should not be expected to look like their namesakes any more than we expect the Hudson River to resemble Henry Hudson or a Louis XIV chair to bear a likeness to that monarch.

Of the various legends which ancient peoples built around the Great Bear constellation, the one in Greek mythology is probably the best known. Briefly, the story concerned a beautiful nymph, Callisto, with whom Jupiter, king of the gods, fell in love. This so aroused the anger and jealousy of his queen, Juno, that she turned Callisto into a bear and drove her into the forest. Callisto had a son Arcas who was a mighty hunter. One day Arcas, not knowing his mother's fate, was about to kill her in her guise of a bear. To avert this tragedy, Jupiter changed Arcas into a bear also and then placed both mother and son in the sky where they would be safe from Juno. And so we have in the heavens the constellations of the Great Bear and the Lesser Bear. Further enraged at being outwitted and at seeing her rival honored in this fashion, Juno conspired with Oceanus, god of the ocean, to have Callisto forbidden ever to bathe in the sea, and so it is that the Great Bear never sets in the ocean as do other constellations, but travels round and round the North Star in a never-ending circle.

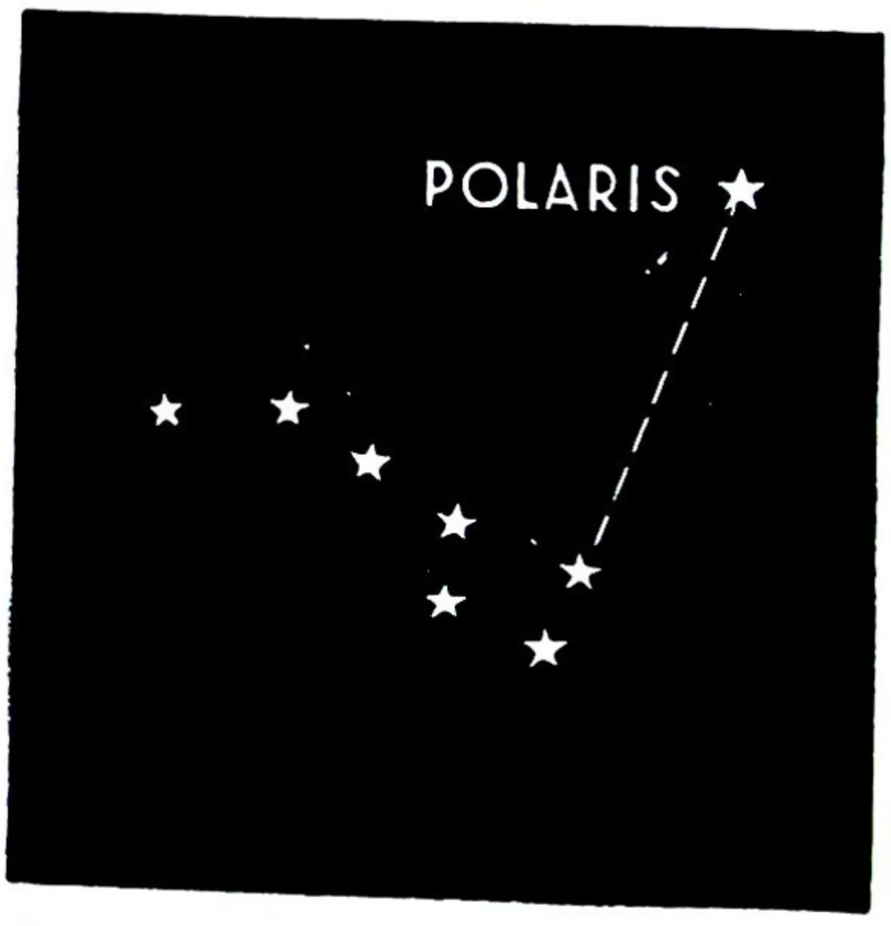


FIG. 16. Relative positions of the Big Dipper and Polaris about 8 P.M. of October 1.

Figure 16 shows how the Big Dipper appears to observers somewhere around the fortieth degree of latitude at about 8 o'clock in the evening during the latter part of September. The reader may wish to use this to check his operation of the northern star chart discussed above.

Polaris

Polaris is so named because at the present time the north end of the earth's axis, or north pole, points almost directly toward it. As will be obvious from Fig. 16 it may be found by following a straight line drawn through the two stars in the Big Dipper which mark the end of the bowl opposite the handle. Although only a second magnitude star, Polaris has few bright stars near it and so is not hard to identify. It is several hundred light years distant. It will be noted in the chart that it forms the end of the handle of the Little Dipper in the Lesser Bear constellation mentioned above.

Cassiopeia

An interesting constellation is Cassiopeia. This group of stars has the shape of a distorted letter W and is sometimes called the Lazy W. It will be found on the opposite side of Polaris from the Great Bear with the top of the W toward the Pole Star. The name Cassiopeia also comes from Greek mythology. This lady was the queen of Ethiopia and the mother of Andromeda. She also had difficulties with Juno, as a result of which she found herself banished to the sky. In Cassiopeia's case it came about because she boasted too much of her beauty and, both as a punishment for her and as a warning to all vain women in the future, she was condemned to hang head downward half the time as she pursues her endless course around Polaris.

Vega and Arcturus

Two first magnitude stars in the northern sky which are not difficult to locate are Vega in the constellation Lyra and Arcturus in Boötes. Vega is visible most of the year but is at its highest point in the heavens during August. It is some 26 light years away and appears to us as the second brightest star in the sky. It will be noted on the chart that if one draws a line to Polaris from the star at the base of the

handle of the Big Dipper, a second line drawn from Polaris at right angles to this one will pass approximately through Vega. Its identification is further aided by the fact that there are two faint stars near it which together with Vega form an equilateral triangle. The entire formation is easily seen with low power binoculars.

Arcturus is below the horizon for northern observers during the winter months and reaches its highest position in the sky during June. It may be located by following the imaginary curved line which would be a continuation of the handle of the Big Dipper. Such a line will pass approximately through Arcturus. The light from this star was used to operate the electric eye, or photo-electric cell, which threw the switch that opened the Chicago World's Fair of 1933. Arcturus was chosen for this honor because of the fact that it is 40 light years distant from the earth. Thus the light that arrived in Chicago just in time to start off the 1933 fair, departed from its native Arcturus during the Chicago fair of 1893.

Sirius and Antares

Since almost all of the United States is located between the 30th and 50th parallels north, the southern hemisphere of the sky offers less of interest to us. Of the stars that are visible in this part of the sky we will mention two. These are both of the first magnitude and should be easily located with the aid of the southern chart during the part of the year when each is visible. The first, Sirius, in the constellation of the Great Dog or Canis Major, is the brightest star in the sky. It will be recalled that earlier in the chapter it was mentioned as having a magnitude of —1.58. It is some nine light years distant from the earth. To observers located around latitude 40 N., it is visible from about December 15 to April 15. The second of the two is Antares. It is red in color and more than 27 million times as large as our sun. Its average density is believed to be less than that of our atmosphere at sea level.

Alpha Centauri

Alpha Centauri, the star already mentioned several times as being the nearest one to the solar system that is visible to the naked eye, is never visible to observers in the United States, as will be evident from the southern sky chart.

Space will not permit us to go further on our trip among the constellations. However, the stars will be available for study for a long time and once the reader has learned his way among the few that have been mentioned, branching out to others should prove interesting and not too difficult.

Systems Beyond Our Galaxy

One group of bodies in space remains to be mentioned. They are located far beyond our galaxy and appear to us as small nebulous bright spots with fuzzy irregular outlines. From their shape and from the fact that they seemed to resemble the nebulae within our galaxy they were named spiral nebulae. Later evidence has shown rather conclusively that actually they are simply other clouds of stars, other galactic systems outside and beyond our own. The nearest ones are almost a million light years away from us and the farthest discovered so far, some 500 million light years distant.

Only a few are visible to the naked eye and of these only one to inhabitants of the northern hemisphere. This one, known as the Andromeda Nebula, may be detected on a clear moonless night as a hazy spot in the constellation of Andromeda. The time of year when it is highest in the sky may be determined from the northern sky chart. Although it is some 900,000 light years away, photographs in recent years have shown it to be made up of separate stars. Its diameter has been calculated to be of the same order of magnitude as that of our own galaxy.

Many more of these outer galactic systems have been photographed (Fig. 17). In some parts of the sky it is possible to get pictures of hundreds of them on one plate. The total number that can be photographed seems to depend only on the power of the telescope in use. It is estimated, for example, that there are several million in the entire sky whose pictures could be taken using the Mt. Wilson 100-inch telescope. How much this figure will be enlarged when the 200-inch instrument is put into use remains to be seen.

Thus it is probable that our galaxy, from a point several million light years away, would present just the sort of appearance that these spiral nebulae do to us. Apparently the only unique thing about our galaxy is that we know it contains at least one star which has a family



FIG. 17. Spiral nebula Messier 51. Photographed with 82-inch reflecting telescope. (Yerkes Observatory and University of Chicago Press.)

of planets, at least one of which supports a system of organisms which possess the peculiar sort of characteristics that we call life. With regard to the other galaxies we have no evidence on the point one way or the other. All this may make man appear a bit insignificant, so to bolster up our ego it should be noted that, unimportant as man may seem in this general cosmic scheme, at least he has had the intelligence and ambition to study and discover these things about his universe.

QUESTIONS

Discussion

1. Express the thickness and diameter of our galaxy in miles.
2. Express the diameter of Pluto's orbit as a fraction of a light year.
3. How can the appearance of the Milky Way be considered as evidence of our galaxy being much broader than it is thick?
4. What would the Milky Way be expected to look like to an observer located on Jupiter? On Polaris? On the nebula in Andromeda?

5. Why are they called " fixed " stars?

6. Assuming that a star of any given magnitude is 2.512 times as bright as one of the next magnitude, exactly how many times brighter would a third magnitude star be than a sixth? A tenth than a sixteenth? A first than a third?

7. Why do stars appear of different colors to us?

8. Explain just why it is that Sirius has a negative magnitude.

9. Could a star that cannot be seen with the naked eye be said to be of zero magnitude as far as non-telescopic observation is concerned? Explain.

10. If any given group of stars were listed in order of increasing magnitude, would they also be in order of increasing temperature? Explain.

11. Find out whether any novae have been discovered within the past year.

12. Distinguish each of the following from the rest: (a) stars, (b) dark nebulae, (c) galactic nebulae, (d) extra-galactic nebulae.

13. Report how many of the following you have been able to locate, using the star charts in the front and back of the book:

(a) Polaris	(b) Cassiopeia	(c) Sirius
(d) Arcturus	(e) Antares	(f) Orion
(g) Vega	(h) The Great Bear	(i) Aldebaran (j) Capella

Multiple Choice

1. Andromeda is a member of (a) both our galaxy and the solar system, (b) neither our galaxy nor the solar system, (c) our galaxy but not the solar system, (d) the solar system but not our galaxy.

2. To an observer on Alpha Centauri, the Milky Way would appear (a) as a spiral confined to a small portion of his sky, (b) the way Andromeda looks to us, (c) about the same as it looks to us, (d) as a huge sphere, (e) as an elliptically shaped cloud.

3. Within 5 light years of the earth there are known to be (a) fewer than 10 stars, (b) about 25 stars, (c) several hundred stars, (d) several thousand stars, (e) about a million stars.

4. The maximum number of stars that are visible to the naked eye at one time on a clear night is about three (a) hundred, (b) thousand, (c) million, (d) billion, (e) trillion.

5. A star of the first magnitude appears to be about — (a) 2½, (b) 5, (c) 6¼, (d) 7½, (e) 10¼ — times as bright as one of the third magnitude.

True–False

1. The star Arcturus is about 40 light years distant from the earth.

2. Many second or third magnitude stars may actually be brighter than some first magnitude stars.

3. It is believed that, viewed from any point in our galaxy, the Big Dipper would appear essentially the same shape that it does to us.

4. In an astronomical time exposure photograph of a section of the sky containing both stars and planetoids, the former will appear sharp and the latter blurred.

5. A straight line drawn through the two stars at the end of the bowl of the Big Dipper opposite the handle, if extended, would pass approximately through Vega.

6. During the summer months, Sirius is visible to observers located in northern latitudes.

7. Astronomers believe that in some star clusters, the stars are so close together that it would be impossible for them to have planets traveling about them in orbits.

8. The term " novae " is applied to a certain group of celestial objects which lie outside our galaxy but have temperatures about like the planets.

9. So-called nebulae which are beyond our galaxy are generally believed by astronomers to be other galaxies.

10. Polaris shines by its own luminosity rather than by reflected light from the sun.

Chapter 5

ASTRONOMICAL MEASUREMENTS

In our discussion of the scientific method in the first chapter, we emphasized particularly the importance of basing conclusions on reliable, carefully collected data. Then in the next three chapters we proceeded to make a number of rather dogmatic statements, quite unaccompanied by experimental data, about such things as the composition of the stars and planets, their sizes and temperatures, and the distances between them. This being the case, if the reader acquired the proper scientific attitude in the course of his perusal of Chapter 1, he perhaps should have been increasingly skeptical as he made his way through Chapters 2, 3, and 4.

It is not our intention, however, to leave these statements in their present unsupported condition. We simply postponed the discussion of the methods of making astronomical measurements until we could talk about all of them at once. Consequently this chapter is given over to the consideration of how such determinations are made, the instruments used, and the reasons we have for considering the results to be accurate and dependable. Since the telescope is by far the most important astronomical instrument we have, it will be discussed first.

The Telescope

Historical Background

The earliest use of irregularly shaped pieces of glass for purposes of magnification is lost in antiquity. Spectacles probably were invented around 1300 A.D., although they seem not to have come into wide use until after the introduction of paper and the invention of printing. By 1600 A.D., it is said, there were opticians practising in most of the larger towns of the continent.

Two other instruments of enormous importance grew out of the development of spectacles, both of them coming into being shortly after 1600. One of these was the microscope, which we use in the observation of very small objects, while the other was the telescope, an instrument that finds its place as an aid in our study of the largest bodies in the universe. The telescope was invented by a Dutch spectacle maker, probably Hans Lippershey, about 1608. Galileo, who frequently has been given credit for its invention, wrote that he heard about the instrument when he visited in Venice in 1609. As soon as he returned to Padua, he built a telescope with a magnifying power of 3 diameters, by placing a convex lens in one end of a lead pipe and a concave lens in the other. Later he built telescopes of greater power, finally making one which magnified some 33 diameters.

The years following this invention were eventful ones. We have already discussed its effect on astronomy in providing the necessary evidence to make possible a choice between the geocentric and heliocentric hypotheses. Almost immediately the telescope also had many practical applications. Its value to navigation was soon recognized. Lancelot Hogben [1] states that Galileo himself was impressed with its military possibilities and "offered the invention consecutively to the Catholic emperor and to the opposing Protestant democracy in letters adapted to the convictions of both parties." The two types of telescopes that are in use today will be considered individually in the next few pages.

The Refracting Telescope

Almost everyone has had some experience with the way in which a convex lens can be used as a burning glass to focus sunlight and set fire to small bits of paper or other combustible material. Since the sun's rays as they normally fall on the paper do not generate enough heat to cause combustion, one may inquire exactly what the lens does to make them so powerful. As Fig. 18 indicates, the sole function of the burning glass is to collect the amount of light that would normally fall on a circular area the size of the lens itself, and bend or *refract* it in such a way that it is concentrated or focussed on a smaller

[1] Hogben, Lancelot, *Science for the Citizen*, p. 123. Alfred A. Knopf, New York: 1938.

area. This focussing or piling up of the sun's rays may produce a temperature high enough to cause the paper to burst into flame.

Now in the refracting or lens type telescope (see Fig. 19) the larger or objective lens performs this same function. It collects light from the star or other body under observation and focusses or concentrates it, producing a small, bright, and inverted image of the body at the point indicated. This small bright image is magnified by the other lens, called the eye-piece. Varying magnifications are achieved by

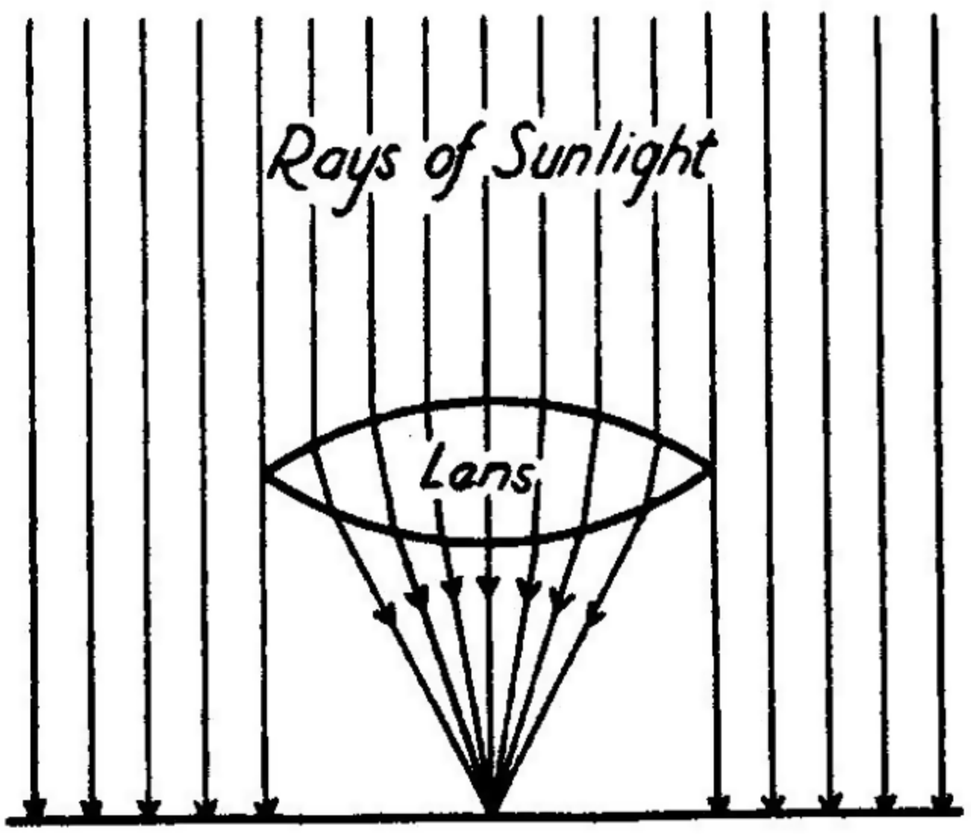


FIG. 18. Diagram illustrating the use of a convex lens as a burning glass.

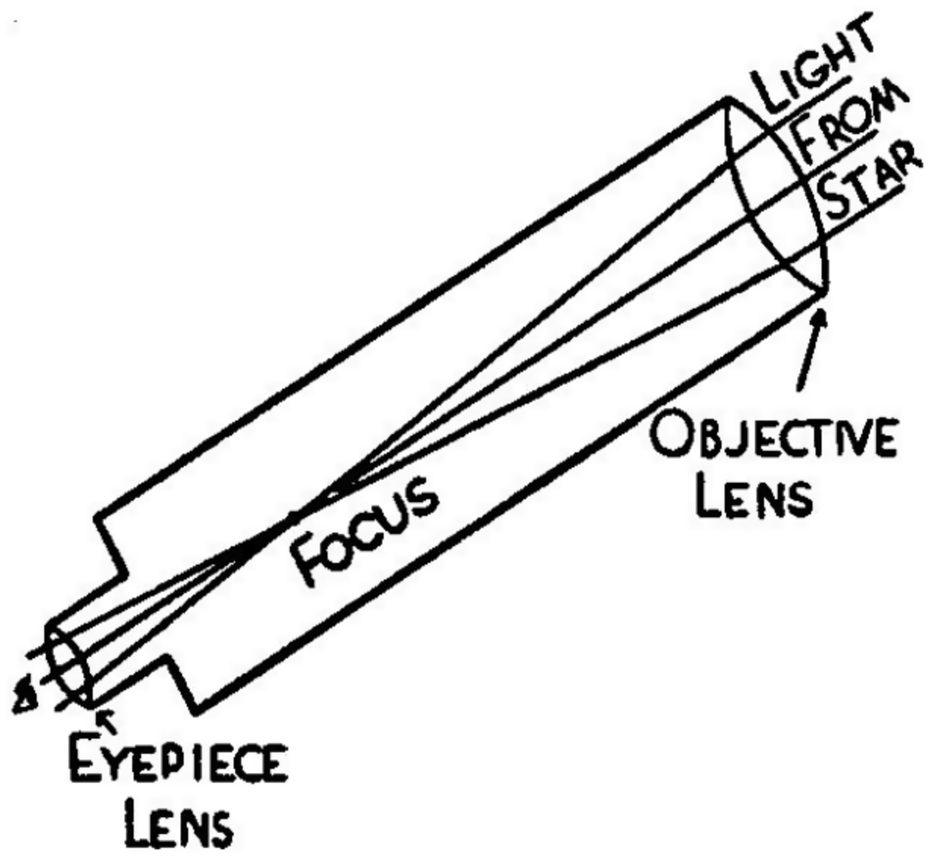


FIG. 19. Diagrammatic representation of the lens system in a refracting telescope.

using different eye-pieces, no magnification whatever being caused by the objective lens. The principle is the same as that employed in binoculars and opera glasses, except that in these instruments provision must be made for producing an upright image.

Those twin adjectives " bigger " and " better," so beloved of us Americans, bear a close relationship to each other in telescopes. That is to say, other things being equal, the bigger the objective lens of a telescope, the more effective instrument it is. In order to see any object it is obvious that sufficient light from it must fall upon the eye. Since the function of the objective lens is to collect the light that would otherwise pass through a circle its size and focus that light, the larger the lens, the more light it will collect and the brighter will be the image. For example, in Fig. 20, a given star would be brighter when viewed through lens *A* than if lens *B* were the objective lens of the telescope, or a star that could not be seen at all using lens *B* might become visible through lens *A* as a result of the additional light col-

lected. Quantitatively, the effectiveness or light-gathering power of lens *A* will be as many times that of lens *B* as the cross-sectional area of *A* is greater than that of *B*, exactly the same as the water-carrying capacities of two pipes are in the same ratio as their cross-sectional areas. However, again like water pipes, lenses are usually described in terms of their diameters rather than their areas; that is, a 6-inch lens is one that is 6 inches in diameter. Thus since the area of a circle depends on the square of the diameter, the light-gathering power of a lens will vary in the same way. Therefore a 6-inch lens theoretically would have a light-gathering power 6^2 or 36 times that of a 1-inch lens. Or we may compare the effectiveness of the Yerkes Observatory telescope, which contains a 40-inch lens, with that of the human eye, the average diameter of which is about one quarter inch. Since the lens in the former has a diameter 160 times that of the latter, it possesses a theoretical effectiveness 160^2 or 25,600 times as great.

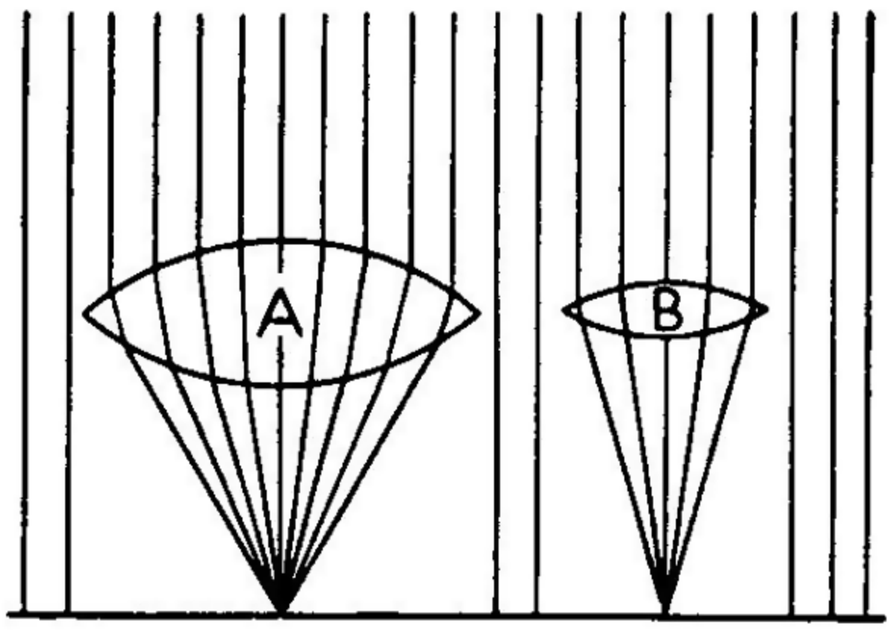


FIG. 20. A comparison of the light-gathering power of two lenses.

The Reflecting Telescope

The first reflecting or mirror type telescope was built by Newton in 1668 and is now one of the prized possessions of the Royal Society of London. In this kind of instrument the light is gathered and focussed by a concave mirror instead of by a lens. Fig. 21 (a) indicates that light is reflected from a flat or plane mirror in such a manner as to make angles *X* and *Y* equal, a reflection law which is the same for all kinds of waves. The same principle holds for such cases as billiard balls being reflected from cushions, rubber balls striking smooth concrete at an angle, or water waves hitting a board. Part (b) of the same figure shows that if the mirror is concave instead of flat, the law of reflection can result in bringing into focus many light rays which

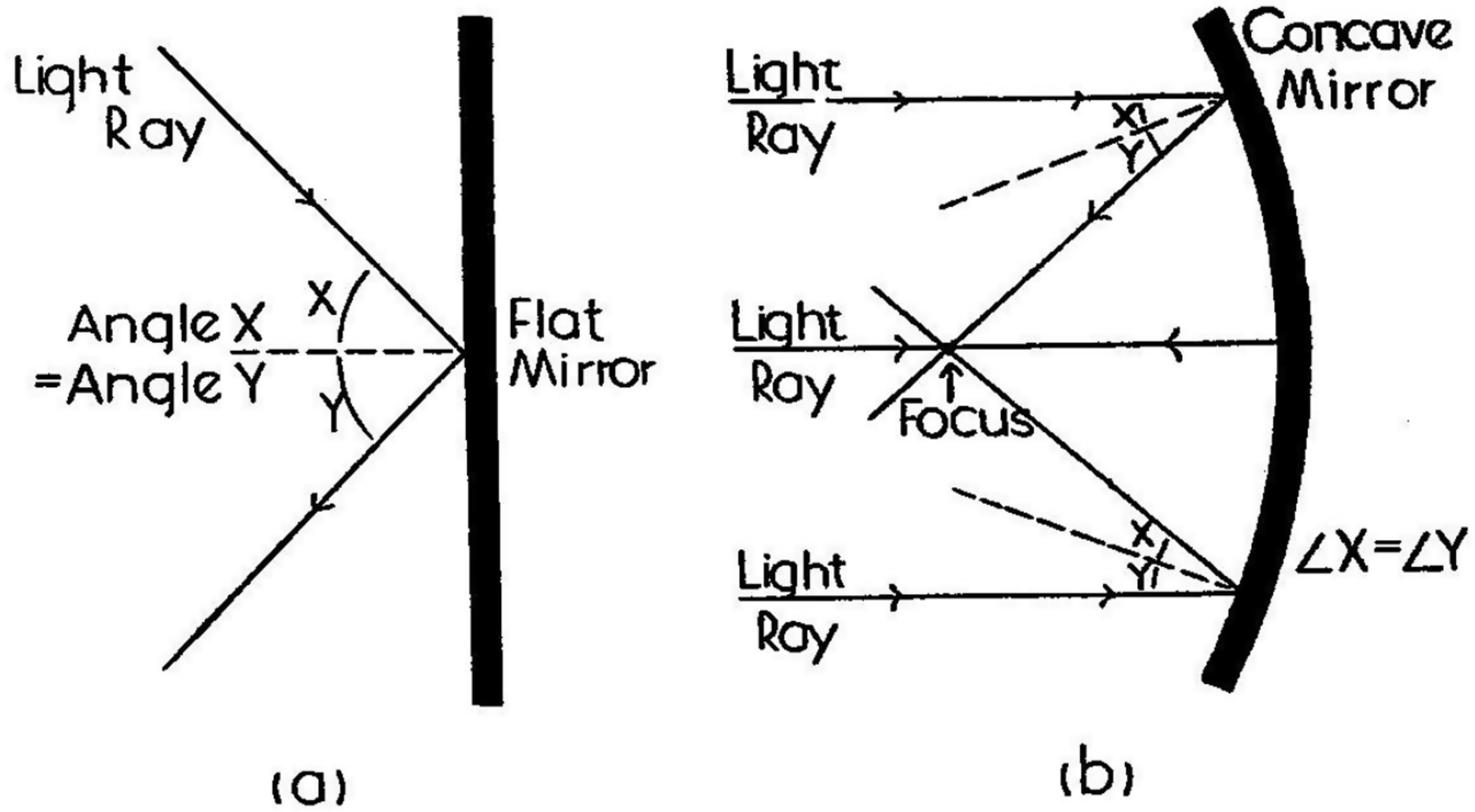


FIG. 21. Diagrams illustrating the law of light reflection for (a) a plane mirror, and (b) a concave mirror.

originally were parallel. The schematic diagram of a reflecting telescope (Fig. 22) indicates that the light from the star or planet under observation is brought to focus by just this means. Thus in this instrument, as in the other, we have a small bright image being produced as a result of many rays of light being brought to a focus. In this case the rays are then deflected off at right angles by means of a small prism or flat mirror (m), and the image is magnified by an eye-piece. It will be noted that here the observer looks in a direction at right angles to the long axis of the telescope, whereas in the refracting instrument he looked along that axis.

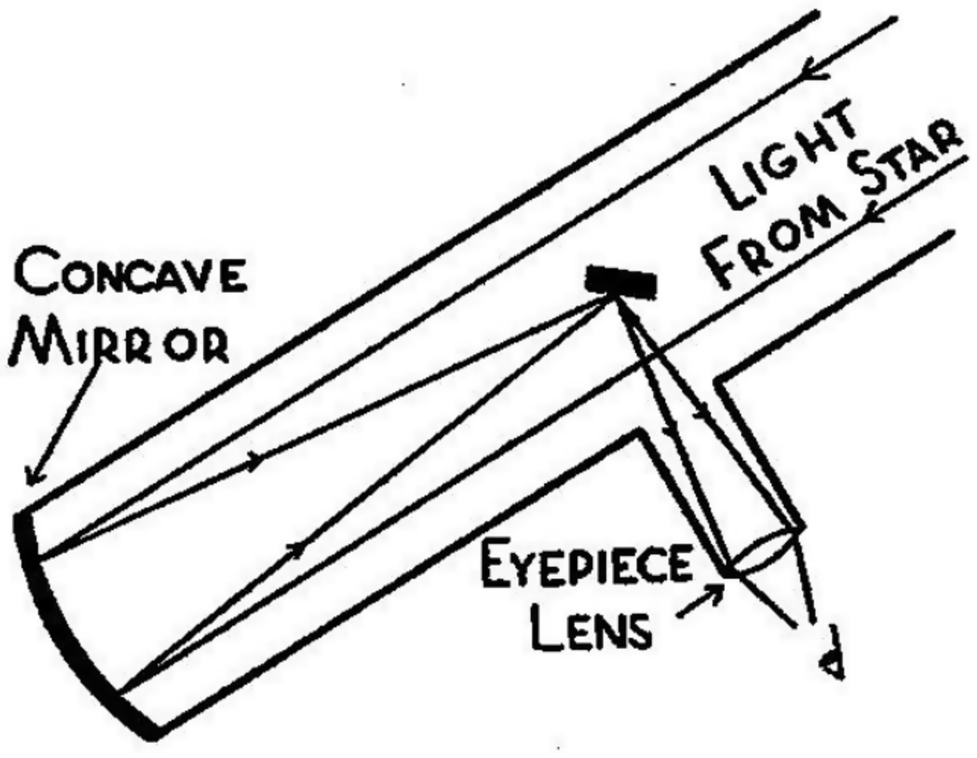


FIG. 22. Diagrammatic representation of a reflecting telescope.

The effectiveness or light-gathering power of this type is expressed by the same relationship as that used in the other case. In other words, the effectiveness of two telescopes will be in the same ratio as the areas of their respective mirrors; that is, as the squares of their diameters. Thus the new 200-inch telescope being set up on Mt. Palomar in California will have a theoretical light gathering power 2^2 or 4 times as great as that of the present 100-inch instrument located on Mt. Wilson.

The mirror type instrument has several advantages over the lens type, the principal advantage being that it can be built in larger sizes. The practical size of a telescope lens is limited by the fact that it has to be flawless throughout and can only be supported around the edge.

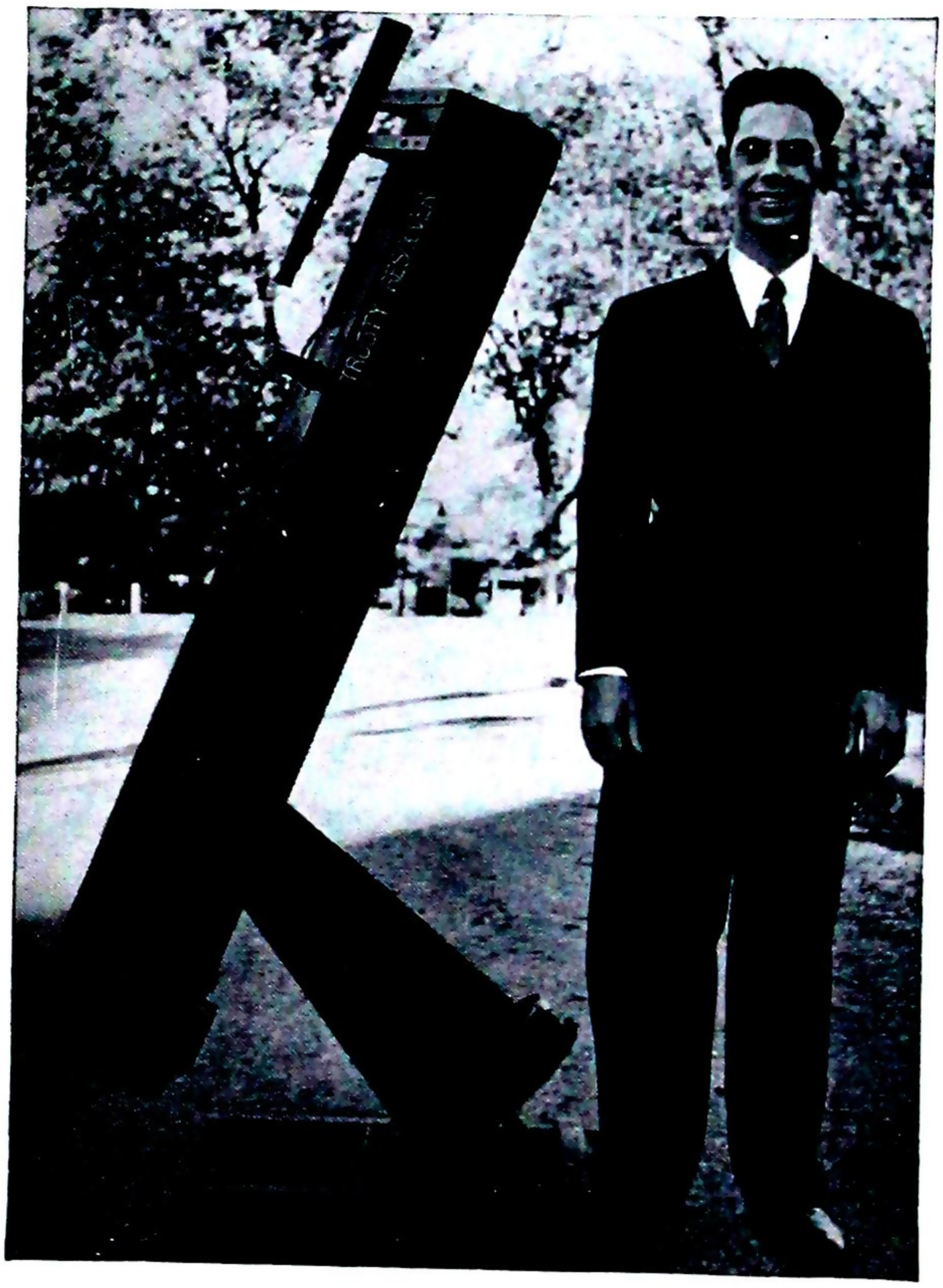


FIG. 23. Reflecting telescope and the builder, Fred Branden. Mirror has a diameter of 6 inches.

Consequently, the 40-inch Yerkes instrument probably represents the maximum that will be attempted in this type. On the other hand, a mirror needs to be perfect only on its surface and can be supported at as many points as desired across its back. Another advantage lies in the fact that certain wavelengths of light which are completely absorbed in going through glass are not so absorbed when the light is reflected from the front surface of a mirror. The reflecting instrument has been of great value to the amateur astronomer because of the

fact that he can grind the mirror himself, whereas a lens might have to be purchased in its finished condition. Thus, reflecting telescopes with mirrors from 6 to 10 inches in diameter can be built comparatively inexpensively. The one shown in Fig. 23 was built by the owner at a total cost of some $20 in addition to the labor involved. The



FIG. 24. Photograph of moon taken using the 6-inch telescope shown in Fig. 23. (Photographed by Samuel Caplan.)

mounting was made from discarded automobile parts. With an instrument of this power a considerable amount of astronomical observation is possible. It has a theoretical effectiveness some 600 times that of the human eye and readily makes visible the rings of Saturn, the four of Jupiter's moons which Galileo first saw, many double stars, and so forth. The photograph of the moon shown in Fig. 24 was taken with this particular instrument, using a camera provided with a ground glass. The eye-piece of the telescope was removed and the camera lens placed as close to the eye-piece tube as possible. The image of the moon was then focussed on the ground glass in the same fashion as that employed for any other picture.

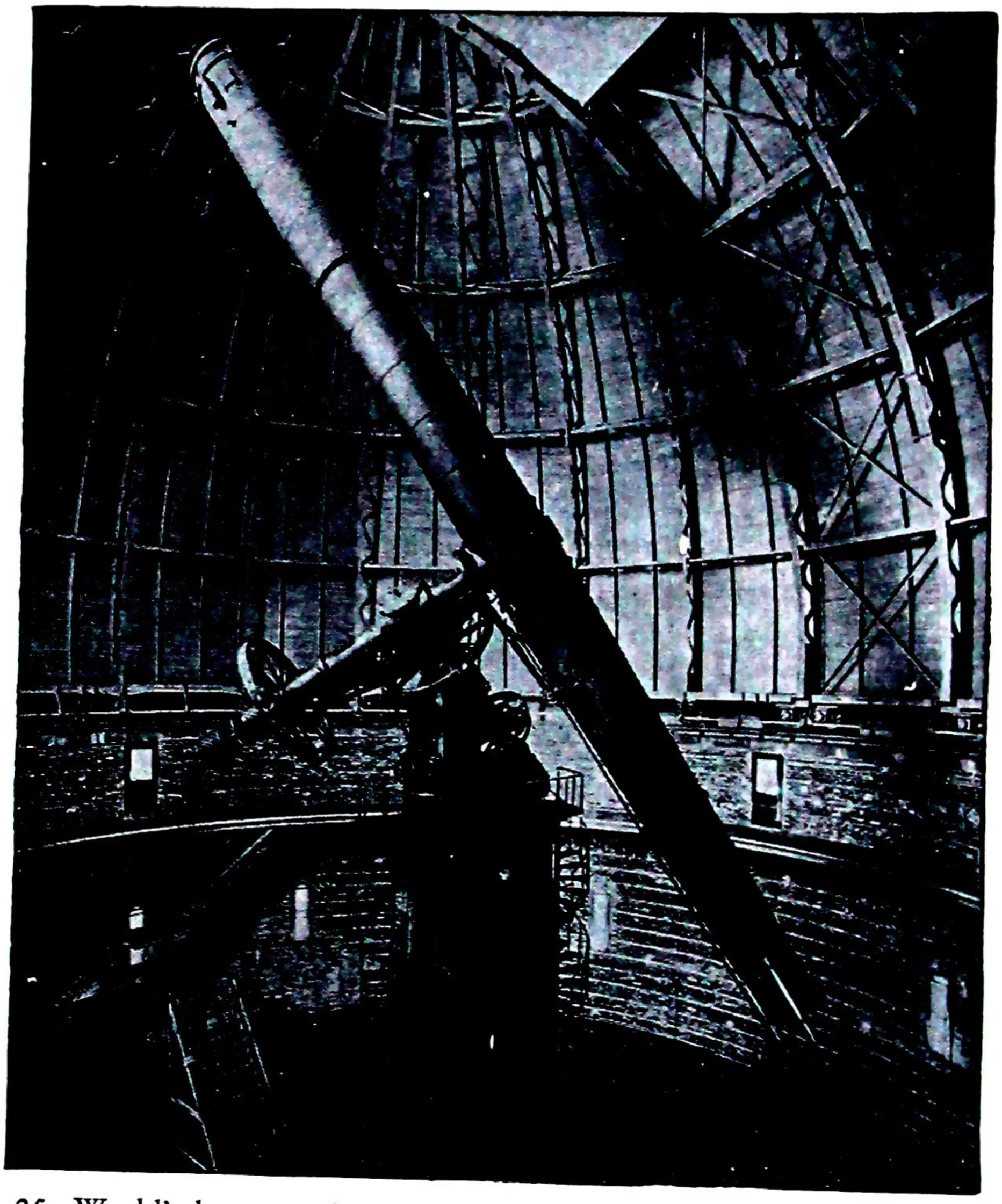


FIG. 25. World's largest refracting telescope, located at Yerkes Observatory. The objective lenses are 40 inches in diameter.

Examples of Modern Instruments

The Yerkes Observatory refracting telescope, shown in Fig. 25, is the largest of its kind in the world. The telescope tube is some 60 feet long and the objective lens has a diameter of 40 inches.

All of the instruments of larger size than this are of the reflecting type. The largest in operation at the present time (June, 1942) is the one at Mt. Wilson Observatory in California. The mirror in this telescope has a diameter of 100 inches, which gives it a light-gathering power some 160,000 times that of the human eye.

The 200-inch reflector (Fig. 26) located at Mt. Palomar in California was to have been completed by 1943. Because of the war, however, construction was stopped from 1941 to 1945, so that now the instrument is not expected to be ready for use before late 1947 or early 1948. The construction and mounting of such a device involve engineering feats of enormous proportions. For example, although the mirror and its mountings weigh over a million pounds the suspension must be such that the cross hairs of the telescope can be set on a single star and follow it across the heavens during an all-night time exposure without causing any blur on the photographic plate. This requires that, as the earth rotates, the instrument must be held within one thirty-six-hundredth of a degree of a single spot in the sky, which is approximately equivalent to keeping a super-machine gun shooting continuously through a 3-foot hoop rolling at a speed of 25 miles per hour 100 miles away. When completed, the telescope should have a light-gathering power over 600,000 times that of the human eye and should open for investigation a sphere in space some eight times that which can now be studied with the Mt. Wilson instrument.

The Determination of the Characteristics of Stars and Planets

Although the telescope is by far the most important astronomical instrument, various other devices are used in conjunction with it to tell us the things we want to know about celestial bodies. The discussion of some of these other instruments follows.

Composition

We have already stated that every element that we know to exist on the other planets and stars has also been found on the earth. Our knowledge concerning the composition of these distant neighbors, which has been largely obtained by the use of an instrument known as a spectroscope, depends on the sort of light they emit under certain conditions.

It was Sir Isaac Newton who discovered that white light is actually a combination of a whole series of colors. He caused white light to enter a three-cornered piece of glass which we would call a prism,

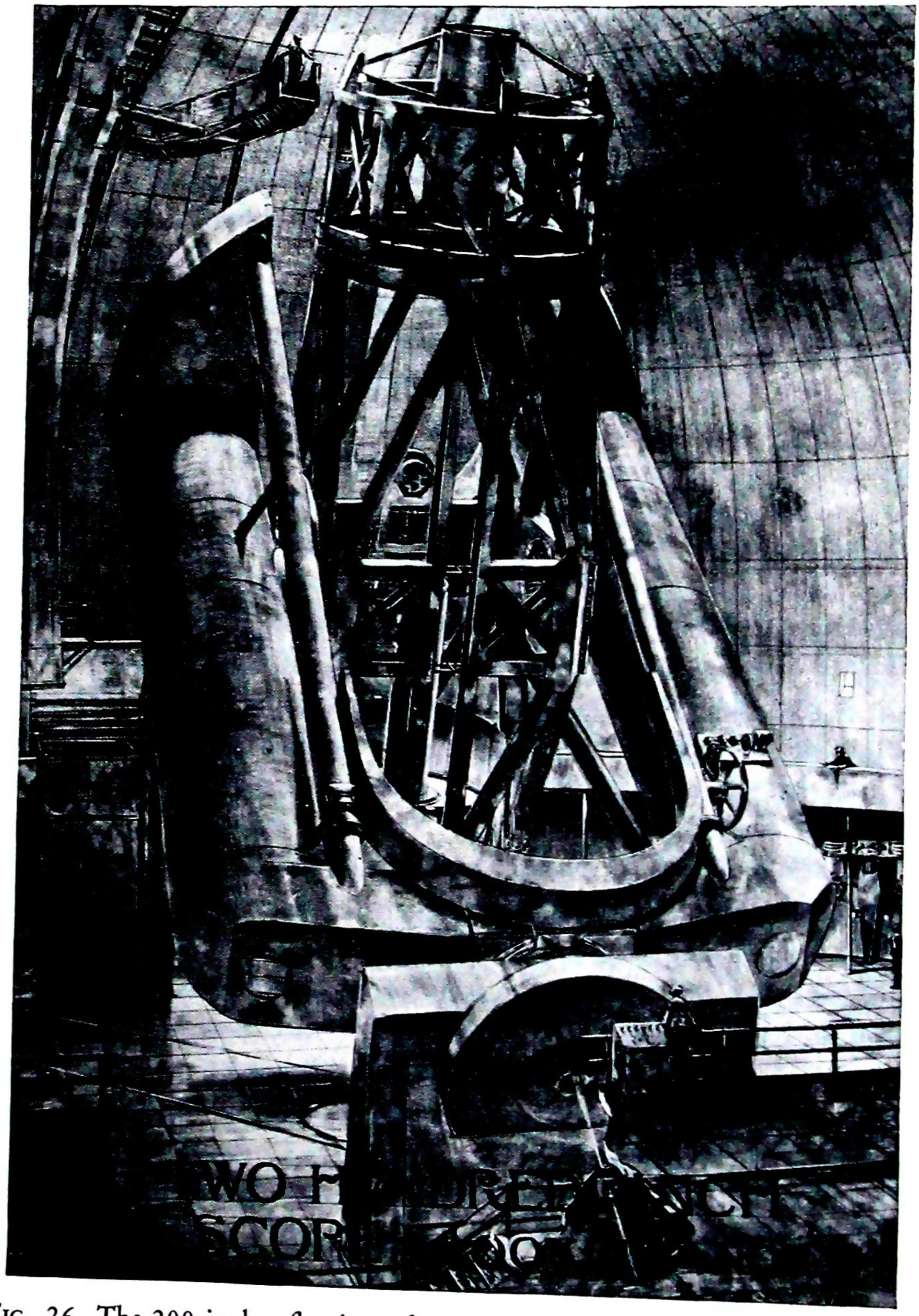


FIG. 26. The 200-inch reflecting telescope located at Mt. Palomar in California. (Courtesy of California Institute of Technology.)

and discovered that the light which emerged on the other side was not white. Instead it consisted of the seven colors of the rainbow as shown in Fig. 27. This type of color sequence is called a *continuous spectrum* because of the fact that the colors merge into each other instead of being sharply separated. All solids when heated to incandescence exhibit spectra of this sort. One of the memory devices frequently suggested for fixing the sequence of the colors in mind is to arrange their initial letters in order beginning with red. When this is done, one can think of the series as being the name of a gentleman called ROY G. BIV.

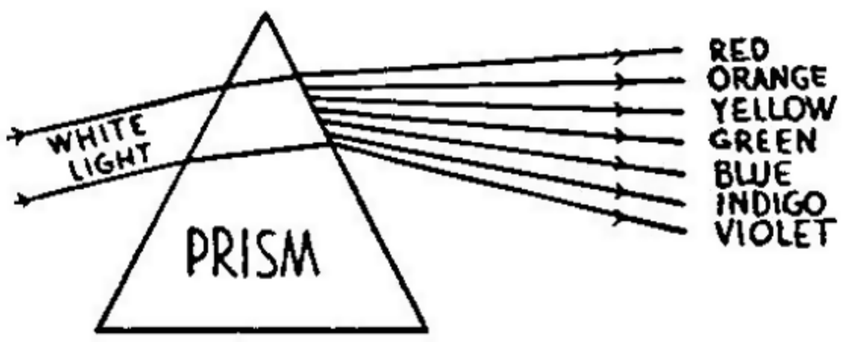


FIG. 27. Diagram illustrating the action of a prism in producing a continuous spectrum.

Now when the prism described above is fixed on a rotating table which is equipped with scales for measuring angles of rotation and telescope tubes to aid in observing the spectrum, the complete device is called a prism spectroscope.

If instead of looking at the white light or incandescent solid mentioned above, we turn a spectroscope toward an incandescent vapor such as that in a neon sign we see a different type of spectrum. In this case there is no continuous band of colors but only a series of bright vertical lines on a black background. These are of various colors depending on their location in the spectrum. It is because several of the brightest lines are in the red region that a neon sign has its characteristic red color. If we shift the spectroscope from the neon sign to a sodium highway light we see only a pair of fine yellow lines very close together. All the rest of the field of vision is black. In like fashion other vapors exhibit their particular patterns under similar conditions. The important feature of this phenomenon, as far as we are concerned with it here, is the fact that every element has its own characteristic set of lines which appear when the substance is in the form of an incandescent vapor. Such a set of lines is called a *bright-line* or *characteristic spectrum*. Thus, if in the spectrum of a source of light as observed through a spectroscope we see the two yellow lines mentioned above, we can be certain that sodium vapor is present in that source. If the sodium lines cannot be seen, we can be certain that sodium either is not present or is there in an extremely low concentration.

One other type of spectrum must be mentioned. A glowing gas has been found to be able to remove or absorb from a continuous spectrum the lines or colors it itself emits. For example, if white light is allowed to pass through sodium vapor and then is observed through a spectroscope, one sees a continuous or rainbow background crossed by two vertical black lines at exactly the location of the yellow lines described above. Such a spectrum is called a *dark-line* or *absorption spectrum*. The colored plate which is the frontispiece of this book gives examples of these various types of spectra. Note that the solar spectrum is of the absorption kind. It happens that this is true of the spectra of almost all stars.

The discovery of helium is a dramatic example of the use of spectra in identifying substances. In 1868 Sir Norman Lockyear found some lines in the spectrum of the sun that were like none known on earth at that time. He attributed them to a hitherto undiscovered element which he named helium in honor of its place of discovery. It was not until 1895 that Sir William Ramsay discovered the existence on earth of this gas with which we are so familiar today. Thus man established the existence of helium at a place 93 million miles away almost 30 years before he found it in his own planetary home.

Distance

A surveyor located on the left side of the river pictured in Fig. 28 would probably determine the distance to the tree on the opposite side in the following manner. He would measure off a so-called base line, *AB*, on his bank of the stream. Then with his surveying instrument located at *A* he would determine the angle *X* between the direction of *AB* and a line from *A* to the tree. In similar fashion he would measure the angle *Y*. Then, knowing two angles and the included side of the triangle *A-Tree-B*, he could easily calculate the other desired dimensions.

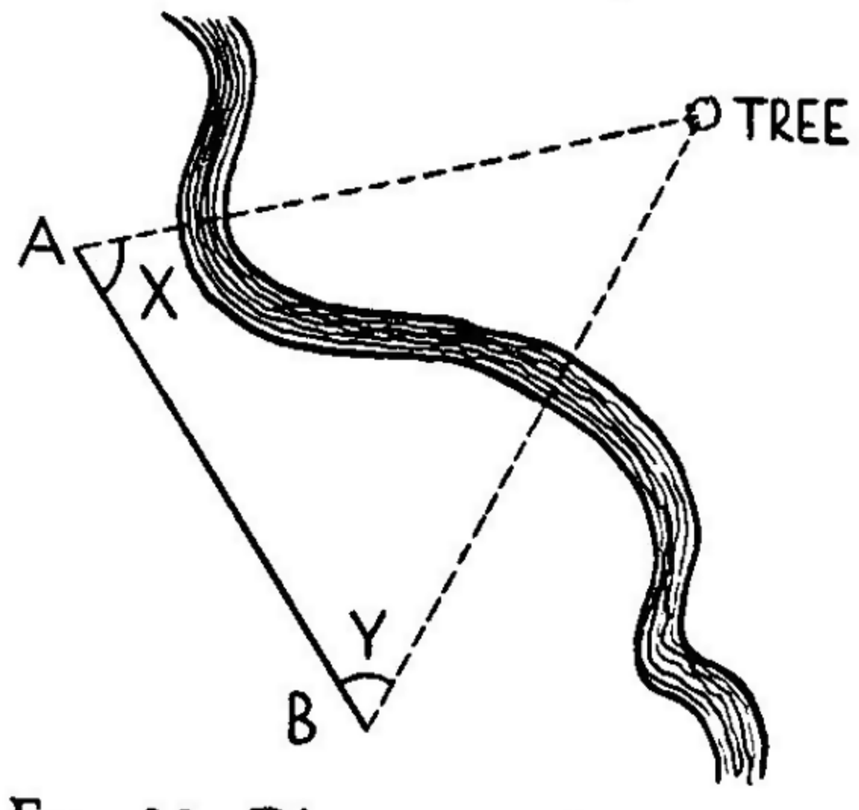


FIG. 28. Diagram illustrating the use of triangulation in the measurement of distance.

A method of measuring astronomical distances, called parallax, is quite similar. In this case the distance

between two observatories, London and Cape Town, for example, may be taken as the base line. Astronomers located at these two points (Fig. 29) take simultaneous observations on the body whose distance from the earth is being measured and proceed to calculate the desired result in much the same manner described in the above example. Sometimes a single astronomer may take all the necessary data by himself by making two sets of observations twelve hours apart. In this case, as a result of the earth's rotation, the base line is the diameter of the earth.

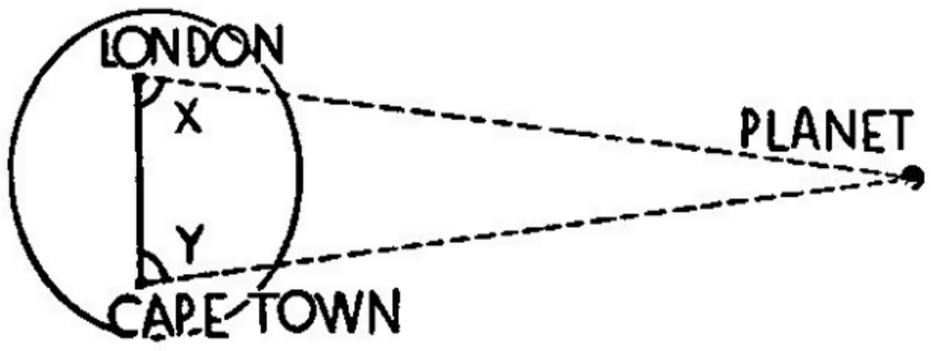


FIG. 29. Diagram illustrating the measurement of astronomical distances using a base line on the earth.

The above procedures give satisfactory results for bodies within the solar system. For stars beyond this group, the angle at the star becomes too small for accurate measurement. For a number of the closer stars a modification of the parallax procedure can be employed. In this case observations are taken at the same observatory at six-month intervals in which case the diameter of the earth's orbit serves as the base line.

Some idea of the accuracy that is possible in modern astronomical measurements may be gained from the fact that even with this immense base line few stars are close enough for the angle subtended to be greater than .2 second. Someone has calculated that such an angle is about the size of one that would be involved in taking measurements on the two sides of a telegraph wire that was four miles away. In the case of Alpha Centauri the angle is about three-fourths of a second.

For stars so distant that none of the above modifications of the parallax technique can be employed, we make use of an indirect method which involves the spectroscope. The method is based on principles which have been well established by laboratory experiment. Such experimentation has shown that the actual brightness of a source of light can be approximately determined if we have a photograph of its spectrum. Quantities such as the widths of the spectral lines and their densities on the photographic film are involved in this calculation. Thus from a star's spectrum we can calculate its actual brightness. Now it is obvious that the farther one is from a source of light,

the less bright that light will appear to be. The quantitative law by which this apparent brightness decreases with distance is well known. Consequently, to get back to the star in question, we can measure its apparent brightness directly, and knowing its actual brightness as determined above can then calculate how far away it is.

In view of the fact that the method is indirect, one has a right to ask how good a basis there is for assuming it to give accurate results. It is considered to be reliable because it can be checked. When it is applied to stars which are close enough to the earth for the more direct parallax methods to be used, results obtained by the two procedures agree. Therefore there would seem to be no reason why it should not give reliable results when used with more distant stellar bodies.

Size

The diameters of the sun, moon, and planets can be measured by parallax methods similar to those described above. The measurements involve determining the angle between telescope settings made on opposite edges of the object. A few stars other than the sun are near enough for direct measurements to be made with the stellar interferometer, invented by Dr. Albert Michelson. This device makes possible the measurement of extremely small angles, the one in use at Mt. Wilson for example, being capable of determining angles down to a thousandth of a second.

For stars more distant than these an indirect method must be employed. It will be obvious that the area of a star could be determined *if* one knew the following two things; namely, the rate at which a square foot of the star was radiating energy and the rate at which the entire side toward the earth was radiating it. A simple division of the second of these quantities by the first would give the number of square feet of the star from which the earth was receiving energy. When one knew this, the entire area and consequently the diameter of the star could be calculated.

Now it has been shown by extensive laboratory experiment that the rate at which unit area of any incandescent surface radiates energy depends *only* on temperature, and the quantitative relationships are well known. Thus, the energy being emitted per square foot of a star can be calculated when we know its temperature, a quantity the

measurement of which will be discussed in the next section. The total rate of energy radiation can be directly observed and the area of the star calculated as described above.

As in the case of star distances, this indirect method is considered to be reliable because it gives correct results when applied to stars which are near enough the earth that direct methods can be used.

Temperature

Approximate star temperatures can be determined from the color of the light emitted by the body. It is a familiar fact that if an object such as a brick or a piece of iron is heated to incandescence and then to still higher temperatures, it first glows a dull red, then becomes bright red, and so on through orange, yellow, white, and bluish white. Thus by matching the color of a star with a color chart for known temperatures, one can attain an approximate value of its temperature. It is said that a good blast furnace man can judge temperatures below 1,500° F. within 45° just by glancing at the color of the flame, so the method is less crude than one might think.

The most common method for determining star temperatures with greater accuracy involves the use of a device called a thermo-couple, an instrument which has also come into wide industrial use in measuring temperatures within various types of furnaces and ovens. This device, involves essentially two strips of dissimilar metals welded together at the ends, making a continuous metallic circuit. Included somewhere in this circuit is a device for measuring small electric cur-rents. Now it is an experimental fact that if the two metallic junc-tions are at different temperatures, an electric current will flow around the circuit. The further fact that the magnitude of the current de-pends on the difference in temperature between the two junctions makes the phenomenon valuable for temperature measurement.

When used industrially to measure furnace temperatures, the hot junction is placed inside the furnace. From the known temperature of the cold junction and the observed magnitude of the electric cur-rent, the temperature of the hot junction and therefore of the furnace can be calculated. In astronomical work, the hot junction is placed at the focus of a telescope. The best thermocouples are capable of detecting temperature differences of less than a millionth of a degree Fahrenheit. The one in use at Mt. Wilson, for example, is made from

bismuth and an alloy of bismuth, and when used at the focus of the 100-inch telescope can detect the heat of a candle 100 miles away.

Velocity

In our discussion of the composition of stars and planets we mentioned the two narrow yellow lines in the spectrum of incandescent sodium. In that case we were assuming the light source and the spectroscope to be stationary with respect to each other. But if the sodium highway light had been on the tail of an airplane which was flying away from our spectroscope at high speed, the spectrum would have looked a little different. The two lines would still have been there; they would have been the same distance apart; they would still have appeared yellow to the eye, but careful measurements would have shown that they had both shifted a little toward the red end of the spectrum. If the light had been approaching us at the same speed, the lines would have been shifted the same distance toward the violet end of the spectrum. In other words, the spectrum of a rapidly moving light source will contain the same number of lines, similarly placed with respect to each other, but with the entire pattern shifted toward the violet or red, depending on whether the light is approaching or receding in relation to the spectroscope. Although the shift is very small and never by any means large enough for the naked eye to notice a change in color, it can easily be measured with modern apparatus and techniques. The fact that the amount of the shift depends on the speed makes it possible to use the effect as a basis for determining velocity.

Thus, if the spectrum of a certain star contains a pattern of lines which appear exactly like the absorption spectrum of neon but with all of them slightly out of place, it is possible to calculate from the amount of displacement the direction and speed of the star with respect to the earth. The reason for such a shift in spectral lines cannot be discussed now but will be taken up in connection with the chapters on Sound and Light. The phenomenon is known as the Doppler Effect, after the man who in 1842 predicted its existence, long before it was actually observed.

Again it is important to note that the method is considered reliable because it gives correct results when used with moving laboratory sources where the speed can be measured independently.

Mass

Any complete discussion of the methods employed in the determination of the masses of stars and planets would lead us into mechanical principles beyond the scope of this book. It can be said, however, that such measurements are based fundamentally on Newton's Law of Gravitation for which there is ample experimental justification, some of which we have considered.

The preceding discussion has not been, of course, a complete and exhaustive consideration of instruments and techniques used in astronomical measurements. It is given rather as a typical set of procedures to illustrate how one may find out some of the things one wishes to know about these celestial objects.

Some methods are more accurate than others; some are more direct than others; but the indirect procedures in general can be checked on some stars or planets by the more direct, and so have their reliability established.

Questions

Discussion

1. Look up some additional information about early spectacle makers and the invention of the telescope.

2. Using a small magnifying glass, try getting on paper an image of a tree or some other object and note whether or not it is inverted.

3. Compare the light-gathering powers of two telescope lenses or mirrors having diameters of 3 inches and 1 inch respectively. Of 6 inches and 40 inches. Of ¼ inch and 100 inches.

4. Give reasons why the largest telescopes are all of the reflecting rather than the refracting type.

5. If you can locate an amateur telescope builder, find out some of the particulars about his instrument. Also try to arrange to view the heavenly bodies through it.

6. After studying the colored page of spectra in the front of the book, what color would you expect incandescent lithium to appear to the naked eye? Hydrogen? Indium?

7. What would you expect to see if you looked through a spectroscope at white light that had passed through hydrogen gas? Through sodium vapor?

8. Describe the discovery of helium.

9. Explain the measurement of distances from the earth to other members of the solar system. To bodies beyond the solar system.

10. Give reasons why astronomers believe the results obtained in the second part of Ques. 9 are reliable.

11. Describe the operation of a thermocouple.

12. Look up some applications of the thermocouple other than those mentioned in this chapter.

13. If the sun were moving rapidly toward the earth how would its spectrum differ from that given in the colored plate?

14. How can the astronomer test the reliability of his method of measuring the velocity of a star with respect to the earth?

15. What measurement described in this chapter would not be possible if a square foot of one material at 8,000° F. radiated energy at quite a different rate than a square foot of some other material at the same temperature?

Multiple Choice

1. Which of the following combinations of instruments would be most valuable for determining the composition of a star or planet? (a) Telescope and thermocouple. (b) Thermocouple and spectroscope. (c) Photoelectric cell and telescope. (d) Spectroscope and telescope. (e) Interferometer and telescope.

2. The mirror of a certain telescope has a diameter of 6 inches; one that would gather in 100 times as much light would have a diameter of — (a) 12, (b) 36, (c) 60, (d) 16⅔, (e) 600 — inches.

3. If white light is observed with a spectroscope after the light has passed through neon gas, which one of the following will be observed? (a) The bright line spectrum of neon. (b) A solid band of red light. (c) A continuous spectrum with all colors present. (d) A continuous or rainbow background crossed by vertical black lines. (e) Vertical white lines on a red background.

4. Of the following, list those which would constitute necessary and sufficient data for the determination of the distance from the earth to a particular star. (a) The star's diameter. (b) Star's actual brightness. (c) Composition of star. (d) Rate at which star radiates energy. (e) Magnitude of star.

5. The spectrum of a star which is rushing toward the earth differs from what it would be if the star were stationary with respect to the earth in that (a) the entire pattern is shifted to one side in the first case, (b) the former has more lines, (c) the former has fewer lines, (d) the former is a bright line spectrum and the latter a dark line, (e) the former is predominantly red and the latter blue.

True–False

1. The mirror in the 200-inch Mt. Palomar telescope performs the same function that the objective lens did in Galileo's telescope.

2. The determination of the total mass of the earth involves the use of Newton's Law of Gravitation.

3. The distance from the earth to a star near Andromeda could not be measured by parallax methods.

4. The color of a star provides sufficient information for a determination of its approximate temperature.

5. A gas absorbs from a continuous spectrum the colors or wavelengths it itself emits when incandescent.

6. The sun is the only star whose size can be determined accurately by parallax methods.

7. All telescopes larger than 40 inches are of the reflecting type.

8. The fact that our largest telescope does not reveal planets traveling around any of the stars outside the solar system is conclusive proof that such planets do not exist.

9. The light-gathering power of a 3-inch telescope is to that of a 4-inch instrument as 9 is to 16.

10. In the thermocouple, an electric current is generated because of a difference in temperature between the two junctions of two metals.

Chapter 6

THE EARTH AS AN ASTRONOMICAL BODY

COMPARED to most of its fellow inhabitants of the universe, the little planet which we call Earth does not appear to be very imposing. In fact, even among the immediate members of its own solar family it does not rank particularly high. Nevertheless, it is our planet, and we have to live on it, so it is of considerable importance to us as human beings. Consequently, we shall investigate it as an astronomical body in much more detail than we did its brother and sister planets.

PHYSICAL CHARACTERISTICS

Athough most people are fairly familiar with the earth's shape, size, and temperature, they might find themselves at a loss if they were asked to show how these quantities are determined. In this chapter we will not confine ourselves merely to stating the data but will attempt also to show something of how one would go about getting them.

Shape

We rather commonly think and speak of the earth as being a sphere that is flattened at the poles, in other words a sphere with an equatorial bulge or " bay window." A somewhat more correct way of describing its shape is to say that the earth is an oblate spheroid.

Its essential sphericity can be shown in a number of ways. There is the fact which everyone learns in his early school days to the effect that the hull of a ship may be invisible while its masts are still in sight. Such a phenomenon could scarcely be true if the ocean were perfectly flat. A more modern as well as a more direct approach to the problem would be to make a series of high-altitude balloon ascensions from a large number of points scattered over the earth's surface. If at the maximum height of every flight, regardless of the portion of the earth

one was above, the horizon appeared to be a circle whose center was directly beneath the balloon, a spherical shape for the planet would be indicated. This has been the appearance of the earth during all balloon ascents that have ever been made. One of the most direct evidences of the spherical shape of the earth is a photograph made by Captain Stevens during one of the flights into the stratosphere by a U. S. Army balloon. The picture, which was taken at the highest point reached during the ascent, clearly shows a curved horizon. Another proof concerns the appearance of the earth's shadow on the moon during a lunar eclipse. This shadow as it moves over the moon's surface is definitely curved.

The existence of the equatorial bulge may be demonstrated with a telescope mounted on an island or ship near the equator. The location chosen must be such that the entire horizon shows water rather than the irregular topography of land meeting the sky. In such a set-up, if the cross hairs of the telescope are set on the eastern horizon and then the instrument is rotated in a horizontal plane they will be found to coincide with the western horizon, but to be above both the northern and southern ones. This can only be true if the earth curves more sharply north and south from the equator than it does east and west.

The flattening at the poles and bulging at the equator are caused by the centrifugal force which results from the earth's rotation. The same phenomenon was mentioned in connection with Jupiter, where it appears to an even greater extent because of that planet's very rapid rate of rotation.

Size and Mass

The earth has a circumference of about 25,000 miles and an average diameter of some 8,000 miles. The polar diameter is about 27 miles less than the diameter at the equator.

The accompanying diagram (Fig. 30) illustrates one method by which the size of the earth can be determined. Assume *A* and *B* to represent observers in constant telephonic communication with each other and located in Columbus, Ohio, and Jacksonville, Florida, respectively. Observer *A* notes a star that is directly over his head. At the same instant *B* notes what star is exactly overhead for his position

in Jacksonville and communicates its name to *A*. Now line *x* is drawn from observer *A* in Columbus vertically upward to the star above him and line *y* from observer *A* to the star over *B* in Jacksonville. Angle No. 1 is the angle between these lines and can easily be determined by *A*. Now since the stars are so far away, lines *y* and *z* are essentially parallel so that angles 1 and 2 are equal. Thus *A* knows the angle which lines drawn vertically downward from Columbus and Jacksonville will make at the center of the earth. In this particular case the angle is about 10 degrees. The shortest distance between these two cities is some 690 miles. Therefore, if an angle of 10 degrees subtends a distance of 690 miles on the earth's surface each degree marks off 69 miles and the distance around the planet is about 360 × 69 or 24,840 miles. Knowing the circumference, one can easily calculate the diameter. This procedure is somewhat like that used by Eratosthenes, some 2,100 years ago when he determined the earth's circumference with an error of only about 3 per cent.

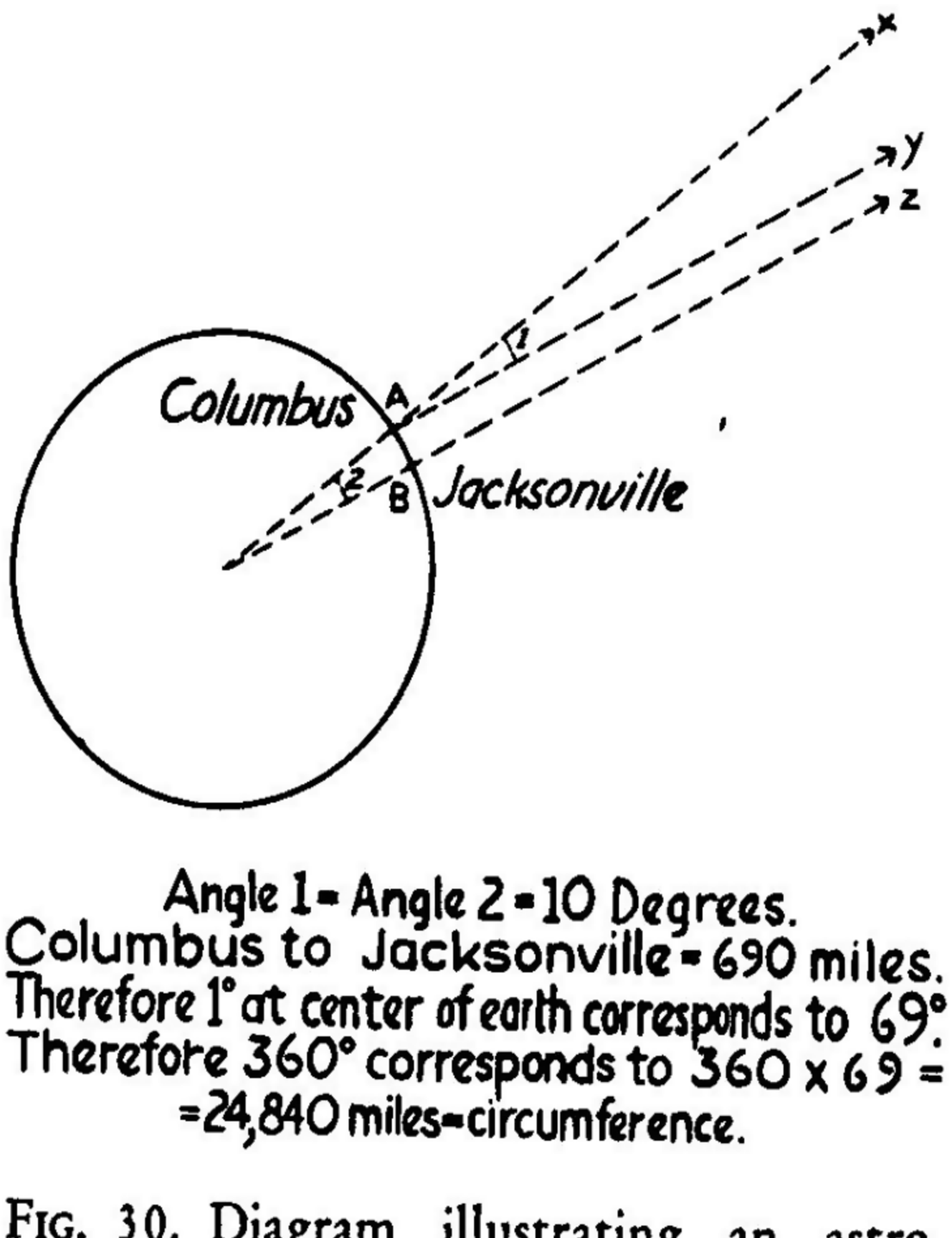


FIG. 30. Diagram illustrating an astronomical method of determining the earth's circumference.

By methods which involve primarily Newton's Law of Gravitation, the mass of the earth has been found to be about 6×10^{21} tons.

Age

The various procedures that have been followed in determining the age of the earth will be discussed in some detail at a later time. At this point it will be sufficient to mention that the method believed to be the most reliable is based on rates of radioactive decay of such materials as radium and uranium. The oldest actual rock deposit investigated by this method is 1,850,000,000 years old. This means that the

earth is at least that old. Certain other considerations have led geologists to adopt a figure of three billion years or more as the minimum age of the earth, while recognizing the possibility that it is much older.

Temperature and Habitability

These two characteristics have been grouped together because the second depends to a considerable extent on the first. In view of the fact that, of the billions of bodies in the universe, the earth is the only one which we know has life upon it, it may be worth while to ask just what set of conditions makes this phenomenon possible.

The first property that the earth has and most other planets do not have is *a comparatively narrow range of temperatures*. All life depends fundamentally on vegetation for food. Animals eat either vegetables or animals, and those animals in turn do the same, until eventually one comes to an animal in the chain that lives on vegetation. Now for vegetation to maintain itself, temperatures in general must not be higher than about 132° F. nor lower than 14° below zero, a range that is extremely small when compared with temperatures on other planets. Earthly temperatures, however, are such that only a small fraction of the planet's surface is ever above or below this range for any great length of time.

Several factors enter into the maintenance of such a narrow temperature spread. The atmospheric blanket, for example, must be of about the correct thickness to let sunlight through and yet act as a tempering agent to help prevent extremes between midsummer and midwinter, or midnight and noon. A different rate of rotation of the earth would affect temperatures since the sun would shine on a particular place for longer or shorter periods than it now does. Bodies of water must be present to provide a tempering effect. The distance of the earth from the sun affects the amount of radiation it gets and consequently its temperatures. Already we have a rather formidable list of requirements just to maintain the proper degree of hotness or coldness. However, all these might be correct and life, as we know it, still be impossible. For in addition to correct temperatures, the atmosphere must have just the right amount of oxygen in it. It must permit the correct amounts of the proper colors of the sun's rays to penetrate it. For example, a population could be destroyed rather quickly

if the ultra-violet rays of the sun were prevented from reaching them. Also carbon has to be present since all organic compounds contain it.

And so one might go on building up a list of conditions, every one of which apparently must be present if this interesting phenomenon which we call life is to exist. One soon begins to feel that it is most remarkable that such a detailed and specific set of properties should have happened to come together on the earth, and that it is quite beyond the bounds of possibility that they should occur anywhere else in the universe.

There is another viewpoint, however, that must not be ignored. We have stated that these conditions must be present if life, *as we know it*, is to exist. This, of course, does not rule out a form or forms of life of a different sort. The fact is that the conditions were here first, and life developed to fit them. Thus it is quite possible that, had somewhat different conditions been present, a different sort of life might have come about, something that may have occurred or may be occurring elsewhere in the universe.

Phenomena which Depend on the Earth's Motions and Positions in Space

There are several phenomena that we, as dwellers on the earth, commonly observe, which are the result of our planet's behavior in space. Some of these occur because of the three types of motion which the earth possesses, and one is due almost entirely to the direction in which its axis happens to point.

Rotation of the Earth on its Axis

We have all grown up with the idea that the earth's rotation on its axis causes the phenomenon which we call day and night, and so we have no trouble accepting it as a fact. As we have seen, however, there was a time when people much preferred to believe that the earth was quite motionless and that the sun, moon, planets and stars moved about it. So we may ask how one would go about proving to such a person that our notion is the correct one.

The fact that such a rotation offers a ready explanation of the equatorial bulge which the earth possesses might be thought of as indirect evidence. But it would scarcely be conclusive in itself. Quite

direct evidence can be obtained by a type of pendulum experiment, devised about 1850 by the French physicist, Leon Foucault. To illustrate the procedure, let us imagine ourselves in a closed freight car which stands on a turntable. Suspended from the ceiling of the car is a long pendulum which consists of a fairly heavy weight hung by a fine wire. The pendulum is started swinging back and forth the long way of the car. So far the car has been stationary, but now suppose that the turntable begins to rotate very slowly. The only direct connection between the pendulum and the car is at the point of suspension so that the only force tending to change the direction of its swing is the twisting action of the wire. If the bob is heavy and the suspension fine, this effect will be negligible and the pendulum will maintain its direction of swing while the car turns under it. Thus when the turntable has rotated through 90 degrees, the pendulum will be swinging across the car instead of in the direction of its long dimension, giving us a means of detecting rotation of the car even if we were unable to see out or otherwise tell that the turntable was in motion. In demonstrating the rotation of the earth, our illustration is analogous to a pendulum suspended over one of the poles. Such a pendulum would appear to change its direction of swing, because of the earth's rotation under it at the rate of 15 degrees an hour, or a complete revolution in 24 hours. As one moves to lower latitudes, the rate of rotation becomes less until at the equator it is zero. The pendulum at the equator would be comparable to a case of the freight car moving along a straight track, that is, with no change in direction. The original Foucault pendulum was set up in the Pantheon in Paris and consisted of a 55-pound weight hung from the ceiling by a wire some 70 yards long.

Another method of demonstrating the earth's rotation consists of taking time exposure photographs of the sky showing star trails. The one shown in Fig. 31 was taken with the camera pointed at Polaris the lens being left open from 8 P.M until 4 A.M. This is exactly the effect that one would expect if the earth rotates. It should be mentioned that taken alone it would not be positive proof of the rotation phenomenon because the same result would occur as far as the photograph is concerned if the earth were stationary and the stars embedded in a huge rotating shell. The short, very bright arc near the center of

the concentric circles is the trail made by Polaris. The fact that it is an arc instead of a point proves that the earth's axis does not point directly toward this so-called Pole Star.



FIG. 31. Star trails in the northern sky, made by pointing camera at Pole Star and leaving lens open from 8 P.M. to 4 A.M. (Photo by Joseph Coburn Smith.)

The apparent rotation of stars at night due to the actual rotation of the earth is illustrated for three times of night in Fig. 32. The Big Dipper and Little Dipper are shown, the times of the night being about correct for October 1.

Revolution of the Earth About the Sun

This motion of the earth determines for us the unit of time which we call the year. Our reasons for believing that all planets possess it have been covered in the general discussion of the solar system. While traveling in its slightly elliptical orbit, the earth maintains an average

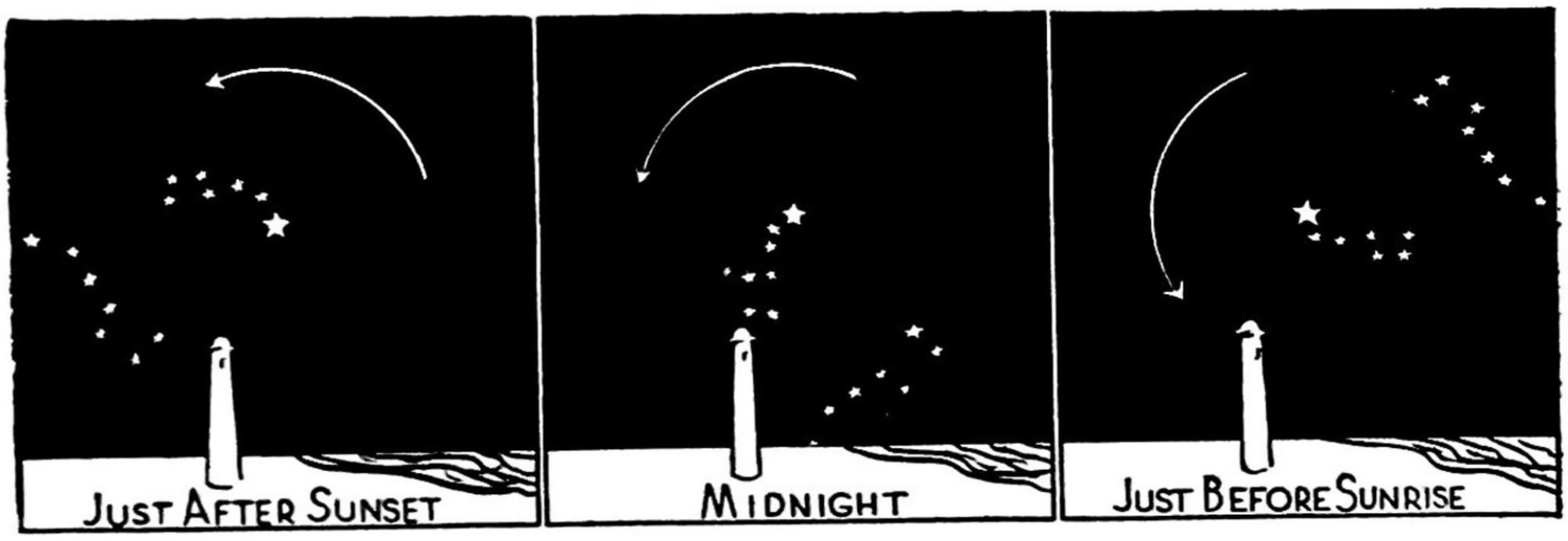


FIG. 32. Apparent rotation of stars due to actual rotation of earth.

distance from the sun of some 93,000,000 miles, being about 3,000,000 miles nearer it in January than in July. Its speed in its orbit averages about 18.5 miles per second.

Precession of the Earth's Axis

Here we are concerned with a somewhat less familiar type of motion of the earth. We may illustrate it by the spinning top shown in Fig. 33. Part (a) shows the top when it is spinning rapidly. It will be noted that the axis about which it spins is straight up and down, that is, perpendicular to the plane or surface on which the top is located. Here there is no precessional motion. But as the top slows down it acquires a new sort of motion which one would commonly describe as " wobbling." As indicated in (b), the axis is no longer perpendicular to the surface but slants, the upper end of it moving in a circle. Such a motion is called *precession*. The axis of the earth precesses or wobbles in just this fashion. We do not notice it because of the length of time required for it to complete one whole " wobble," a matter of some 26,000 years. This means that the North Pole does not always point at the same star. When the cycle has been half completed, about 13,000 years from now, the Pole Star will not be Polaris whose name means " Pole Star," but Vega, all of which should be very

confusing to future students of astronomy. This precessional motion is a direct result of the earth's having an equatorial bulge. This causes an abnormal concentration of mass around the equator, which results in an unequal gravitational attraction of the sun and moon for different parts of the earth.

The three sorts of motion just described for the earth are the same three that would be possessed by a slowly spinning top located on a merry-go-round. Such a top revolves about the music box in the center; it rotates about its own axis; and at the same time the axis " wobbles " or precesses.

Inclination of the Earth's Axis

The reader may have been puzzled at some time as to why globes are always constructed with the axis tilted instead of vertical. At first thought vertical might not seem to have much meaning out in space. It has meaning, however, if we define it with respect to some line or surface. In the case mentioned previously we said the axis of the top when it was spinning rapidly was perpendicular to the surface on which the top was located and was not perpendicular when it began to wobble or precess. Similarly we may define the direction of the earth's axis with respect to the surface or plane that might be thought of as including the orbit of the earth. On this basis, the axis of the earth is not straight up and down but inclined to the vertical at an angle of 23.5 degrees as indicated in Fig. 34. The plane of the earth's orbit is known as the *ecliptic* and the 23.5-degree angle of inclination by the rhythmic title of the *obliquity of the ecliptic*. As a result of this angle of tilt, certain familiar lines on the earth's surface are defined. The Tropic of Cancer is at a latitude of 23.5° N., and the Tropic of

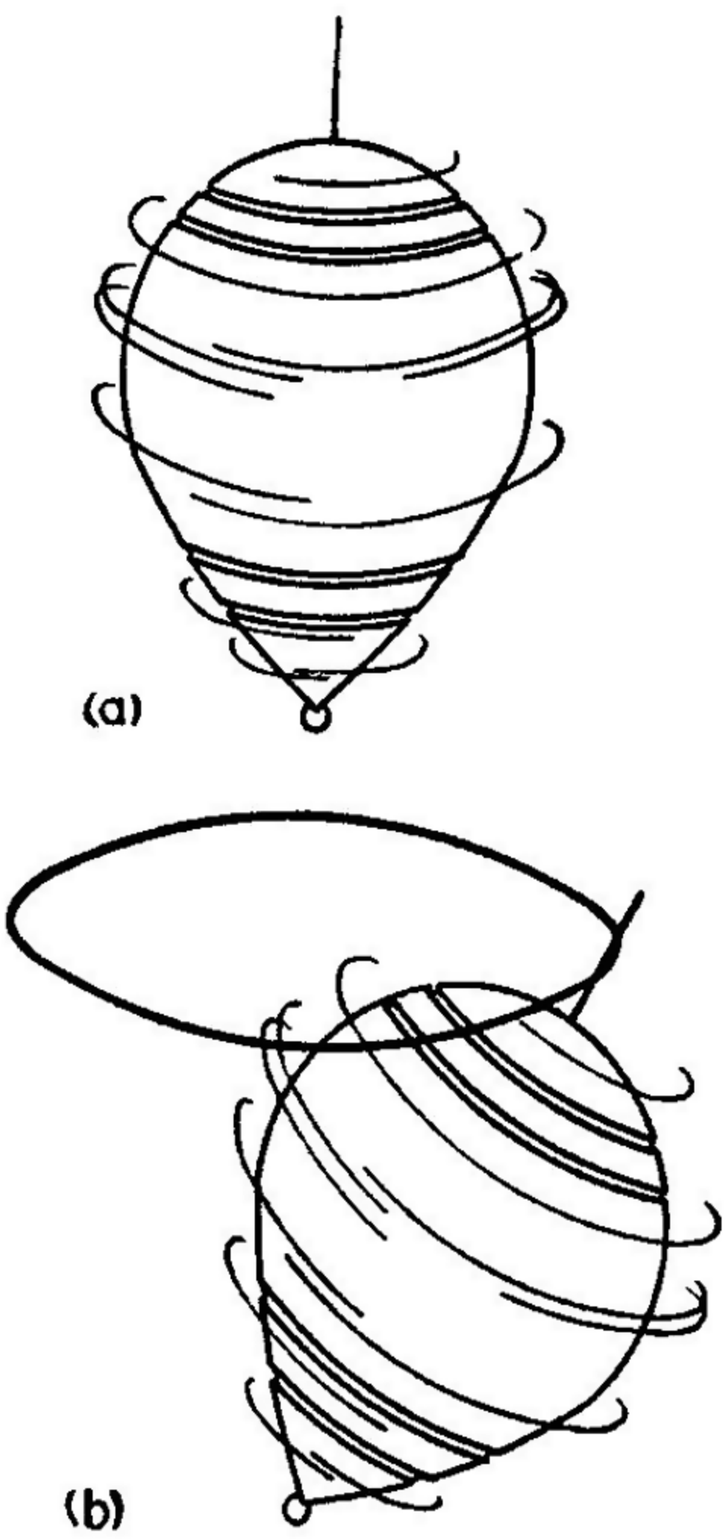


FIG. 33. (a) Top spinning without precession. (b) Top spinning and precessing.

Capricorn the same number of degrees south of the equator. Points 23.5 degrees from the North Pole are on the Arctic Circle and have a latitude of 66.5° N. The Antarctic Circle is similarly defined with respect to the South Pole.

The principal significance to us of the inclination of the earth's axis lies in the fact that it is almost the sole cause of the sequence of seasons. This will be shown by a study of the diagram in Fig. 35. Here the earth is pictured at four times during the year. At each of these times we shall consider temperatures and comparative length of day and night.

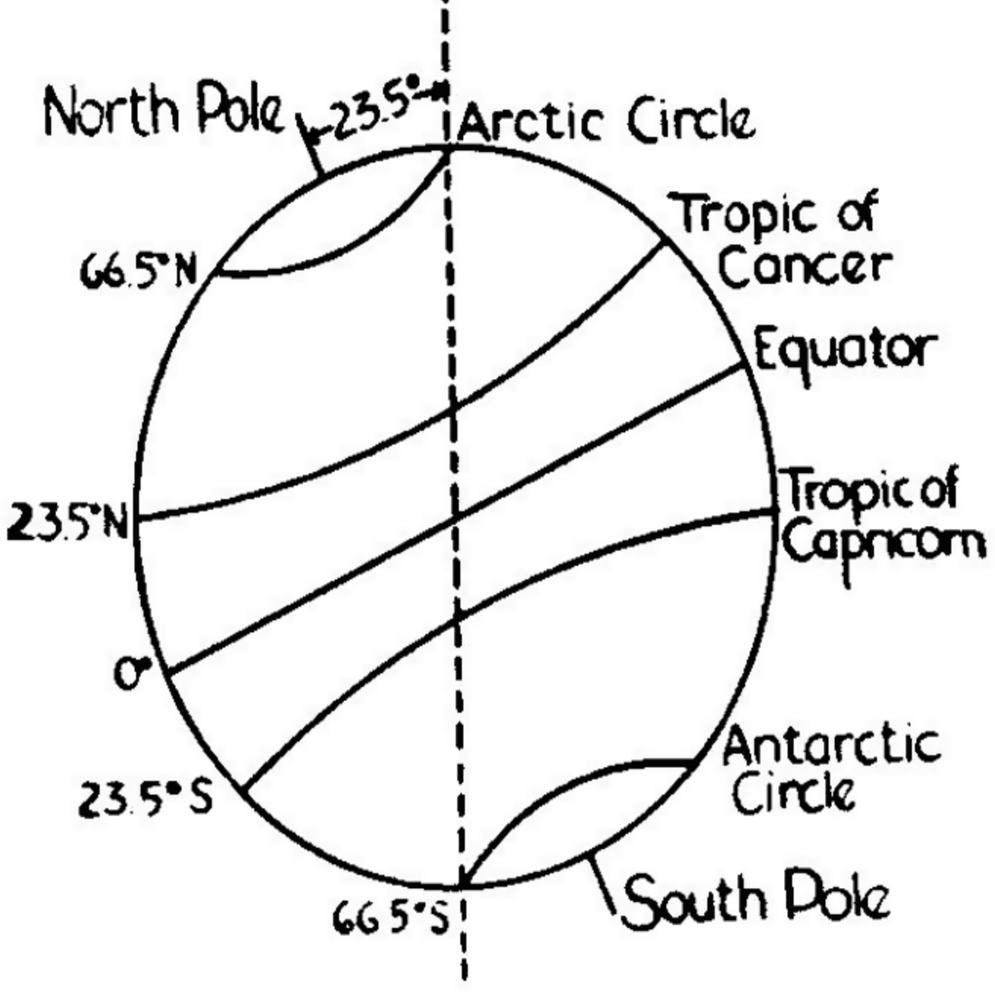


FIG. 34. Location of Tropics, Arctic and Antarctic Circles.

On or about June 21 the sun's rays fall perpendicularly on the Tropic of Cancer and least directly on points in the southern hemisphere. Therefore at this time the northern hemisphere might be expected to have its warmest weather. Actually there is a lag caused by the time necessary to heat up the atmosphere and earth's surface so that maximum temperatures in the northern hemisphere occur a little after June 21. At the same time the southern half of the earth is having winter. By about December 22, the North Pole is tilted away from the sun instead of toward it and temperature conditions in the two hemispheres are the reverse of what they were on June 21. Halfway between these two dates, that is about September 23 and March 21, the equator is receiving the direct rays of the sun.

A similar series of changes occurs with regard to comparative lengths of day and night in the two hemispheres. Points on the equator have equal day and night the year round. As one goes north from the equator on June 21, the number of consecutive hours of daylight increases until at the Arctic Circle the sun shines continuously for approximately 24 hours. At the North Pole the period of sunshine covers the 6-month interval from March 21 to September 23. From

the equator south, on June 21, days are shorter and nights longer until at the Antarctic Circle there is a full 24 hours of darkness, and 6 months without the sun at the South Pole. It is obvious that on December 22 these conditions will be reversed with respect to hemispheres, places north of the equator having their shortest day and points south the longest. Day and night will be equal for all points on the

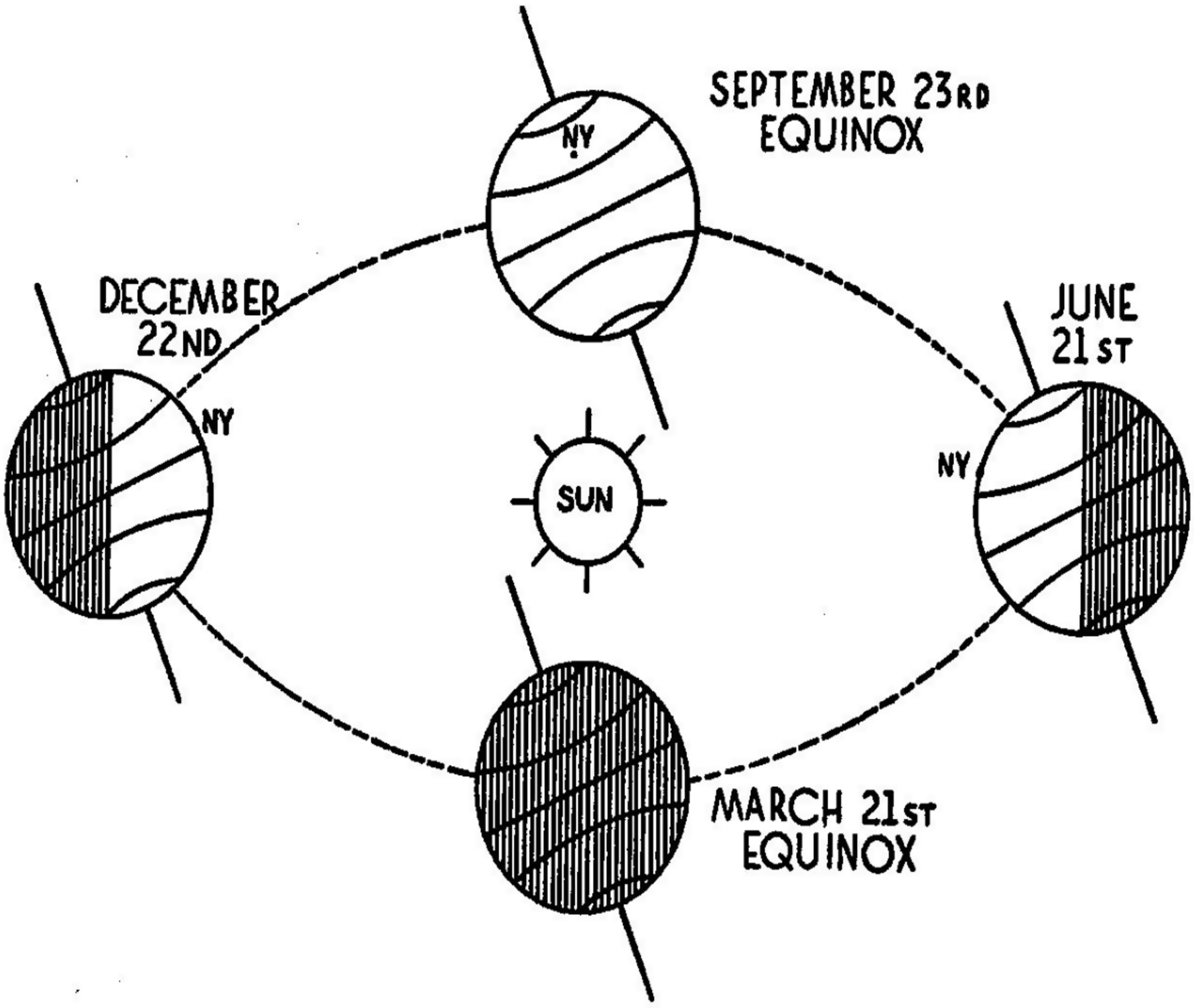


FIG. 35. Diagram illustrating the relationship of the inclination of the earth's axis to seasons and relative lengths of day and night.

earth's surface about September 23 and March 21, from which these times of the year get their name of equinoxes.

The way the so-called midnight sun appears to move across the northern horizon when observed from a point near Circle City, Alaska, is shown in Fig. 36. The photographs were taken at 15-minute intervals.

These varying conditions of temperature and of relative lengths of daylight and darkness would not exist if the axis of the earth were perpendicular to the plane of the planet's orbit. In that case the sun's rays would strike any given point at the same angle throughout the year, while day and night would be of equal length at all places all the time.



FIG. 36. The midnight sun apparently moving across the northern horizon as the earth rotates. The photographs were taken at 15-minute intervals from a point near Circle City, Alaska. (Courtesy Co-op. Photo Shop, Fairbanks, Alaska.)

THE MOON

Physical Characteristics

The earth's only satellite is a cold, dead sphere with a diameter of some 2,160 miles, about one-fourth that of the earth itself. Its average distance from the earth is about 240,000 miles, close enough so that the best telescopes now in use enable us to see objects on it of sizes from 300 feet up. As Fig. 37 indicates the moon's topography includes great rolling plains and vast rugged mountain ranges. Some of the latter extend as much as four and five miles above the surrounding lowlands. The thousands of craters, varying in diameter from a fraction of a mile to over 60 miles, form one of the most striking features of the lunar landscape. Various hypotheses have been advanced to account for their formation. Some astronomers suggest that they were caused by the impact of meteors on the moon's surface. Against this theory it is argued that such an event would have resulted in the craters being scattered more or less uniformly over the moon's surface, whereas actually they are almost entirely confined to the mountainous regions. Another possibility is that they represent the dead craters of ancient volcanoes, but no direct evidence exists for either theory.

It seems quite certain that the moon has no atmosphere. In the first place, because of its small mass the satellite's gravitational attraction is probably not sufficient for it to hold atmospheric gases. Then

the extremely sharp and jagged appearance of mountain peaks suggests that no atmospheric weathering has been occurring. This is further substantiated by the distinct, sharp appearance of the edge of the moon, since the presence of gases should make it look hazy to us.



FIG. 37. The moon, age 10.35 days. Taken with 36-inch reflecting telescope. (Photo by J. H. Moore & J. F. Chappell, Lick Observatory.)

Spectroscopic evidence indicates that there is no water on the moon. Thus, the absence of both atmosphere and oceans to act as tempering agents would lead us to expect the great extremes of temperature which are actually found to exist. Temperature measurements show that changes from 230° F. at midday on the moon to

—110° F. at midnight may occur. It has been suggested previously that very long days and nights would make for extremes of temperature, so that the fact that a day on the moon is over 29 times as long as a day on earth tends to increase the temperature range.

The moon possesses two types of motion. In its revolution about the earth, it travels in an orbit that is roughly elliptical and has a radius of some 240,000 miles. To complete this circuit requires between 29 and 30 of our days, the period of time that we call a lunar month. The plane of the moon's orbit is inclined to that of the earth at an angle of about 5 degrees, a circumstance that has an important bearing on the phenomena of lunar and solar eclipses as well as phases of the moon.

In addition to its motion of revolution, the moon rotates on its axis. As an illustration of the somewhat peculiar manner in which this rotational motion occurs, let us call the reader's attention to one of the wooden horses on a merry-go-round. Note that as this animal travels in its circular path it always keeps the same side toward the organ in the center. Consequently an employee standing beside the organ would see but one side of the horse and so might be tempted to say that the animal does not rotate. An individual standing out in the park, however, gets the correct view of the situation. To him it is evident that the horse does rotate since he sees all sides of it. As a matter of fact, the interval from the time he sees one side of the animal until he sees the same side again is exactly the same as that required for one revolution about the organ. In other words, the horse's period of rotation is the same as its period of revolution. The relationship of the moon to the earth is like that of the horse to the organ, in that the satellite always presents the same face to us. Actually man has seen about five-ninths of the moon's surface because of the greater size of the earth.

The fact that the moon's period of rotation equals its period of revolution about the earth means that any given point on it is in constant sunlight for over 14 consecutive days and then in darkness for the same length of time.

Phenomena Involving the Motion and Position of the Moon in Space

The moon has much less effect on our earthly affairs than is often supposed. There seems to be no scientific evidence, for instance, to indicate that it is of the slightest importance in the proper care and

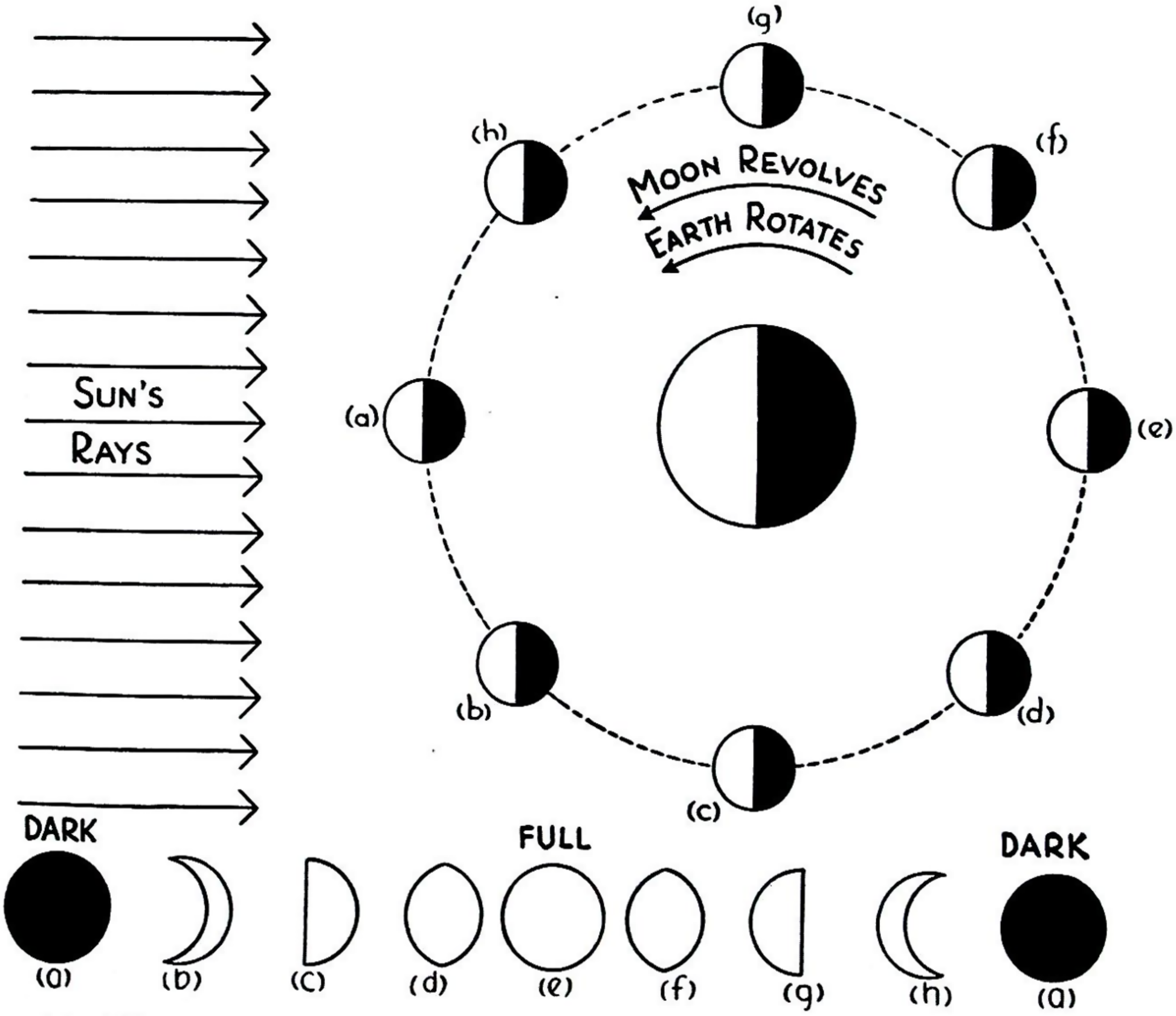


FIG. 38. Diagram illustrating the reason for the occurrence of phases of the moon. (After Menzel.)

feeding of the growing potato. It apparently has little or nothing to do with producing our weather. In fact, aside from its romantic aspects and its commercial value to popular song writers, its influences on us are only about four in number, namely, to provide us with a natural unit of time, which we no longer use; to give us the various phases of the moon to look at; to be a party to the production of solar and lunar eclipses; and to cause tides on the earth.

The Lunar Month. — The revolution of the moon about the earth does provide a natural unit of time, one which is called the lunar

month. This unit was of considerable importance in certain early calendars at a time when it was thought necessary to begin each month with a new moon. The practice was later abandoned because of complications arising out of the fact that the lunar month neither contains an exact, whole number of days, nor is contained an exact, whole number of times in a solar year. Consequently since the time of Julius Caesar, years have contained just 12 months with the new moons falling where they may. The lunar month's place in the calendar will be discussed in greater detail in a later chapter.

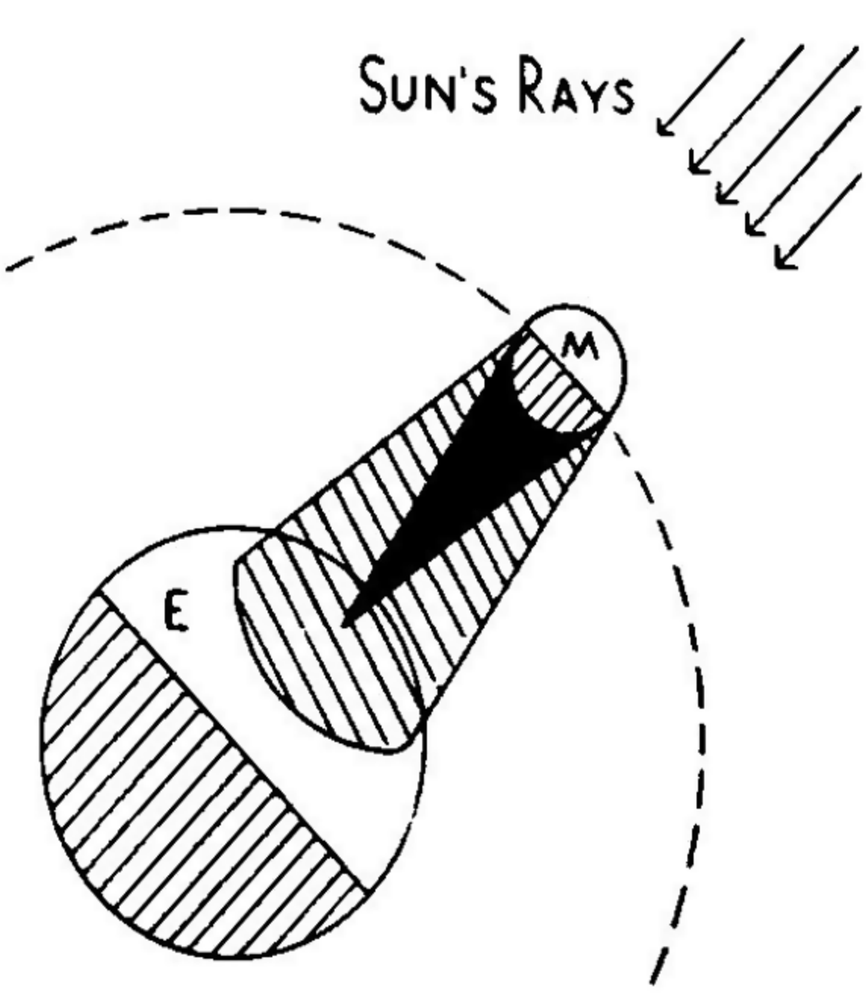


FIG. 39. Relative positions of sun, moon, and earth during a solar eclipse.

The Moon's Phases. — Phases of the moon constitute a phenomenon the cause of which is frequently not understood. To see why it occurs let us trace the moon in Fig. 38 through one complete revolution about the earth. First, recall that the moon, like the planets, is not luminous of itself. It is cold and dead and becomes visible only as it reflects the sun's rays. In the diagram, sunlight is seen streaming in from the left, lighting up the left-hand sides of the earth and moon. Now consider the moon first in position (a), remembering that we are looking at it from the earth. Obviously, the only part of the moon that we can see is the unlighted half. So we speak of the " dark of the moon " or " new moon," not because something is between us and the moon, but merely because there is no sunlight shining on the side toward us. By the time it gets to (b), we can see a larger amount of the lighted half, so that it is past being a " new moon." At (c), half of the side toward us is reflecting sunlight and we see a " half moon." At (d), it is more than half full. By the time it gets to (e), the entire lighted side of the moon is toward us and we see a " full moon." At first thought it may appear from the diagram that at (e) the moon should be in the earth's shadow. But remember that the moon's orbit is inclined to the earth's at a small angle so that in the figure the full moon should be either a little above the paper or a little below it. From position (e),

through (f), (g), (h), and back to (a) again, it reverses the above sequence. Thus, our view of the moon varies according to our position relative to the half of the satellite on which sunlight falls and, as we shall see, in no way resembles the eclipse phenomena with which it is frequently confused. The way the moon looks to us at points (a) to (h), inclusive, is shown at the bottom of the figure.

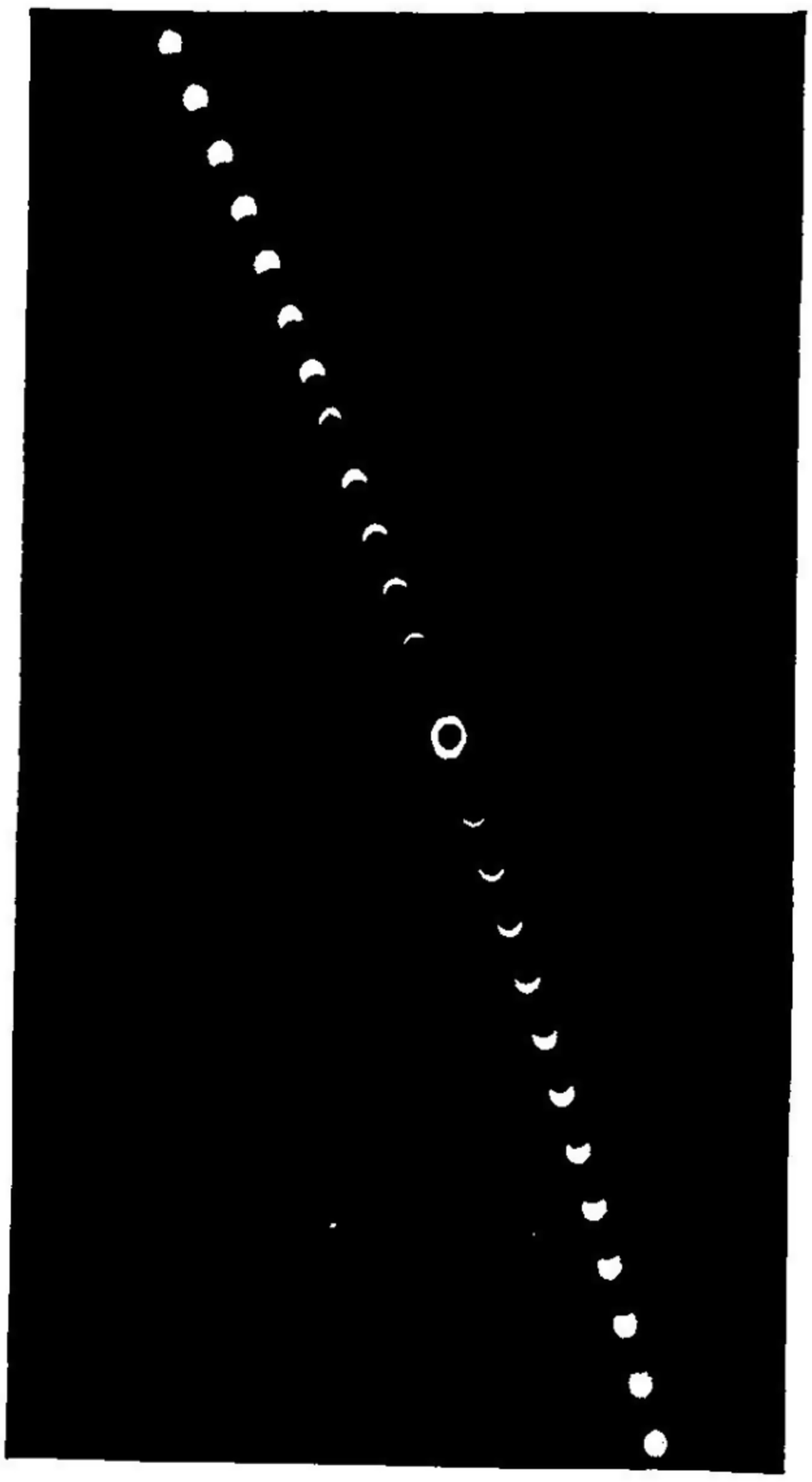


FIG. 40. Series of photographs showing progress of solar eclipse of June 8, 1937. Pictures were taken at 5-minute intervals. (National Geographic Society — U. S. Navy Eclipse Expedition, 1937.)

Solar and Lunar Eclipses. — As a result of the combined revolutions of the earth about the sun and the moon about the earth, it occasionally happens that the three bodies get exactly lined up. If the arrangement is such that the moon is between the earth and sun, the phenomenon that we call an eclipse of the sun occurs. At such a time, the image of the sun is completely blotted out for a small circular area of the earth's surface (Fig. 39). Surrounding this region of total eclipse is a ring-shaped area for which the sun is partly covered, called the region of partial eclipse. Since the distance of the moon from the earth varies slightly from time to time, it occasionally is so great at the time of a solar eclipse that there is no area of totality. In such a case a thin ring of the sun is visible entirely around the moon and the eclipse is said to be an "annular" one. The maximum area of totality that can possibly occur is such that the eclipse lasts a total of between 7 and 8 minutes. An average eclipse of the sun lasts for 3 minutes or less. The complete progress of a particular eclipse is shown in the series of photographs in Fig. 40. The individual pictures were taken at intervals of 5 minutes. A view of the sun's corona taken during the same eclipse is given in Fig. 41.

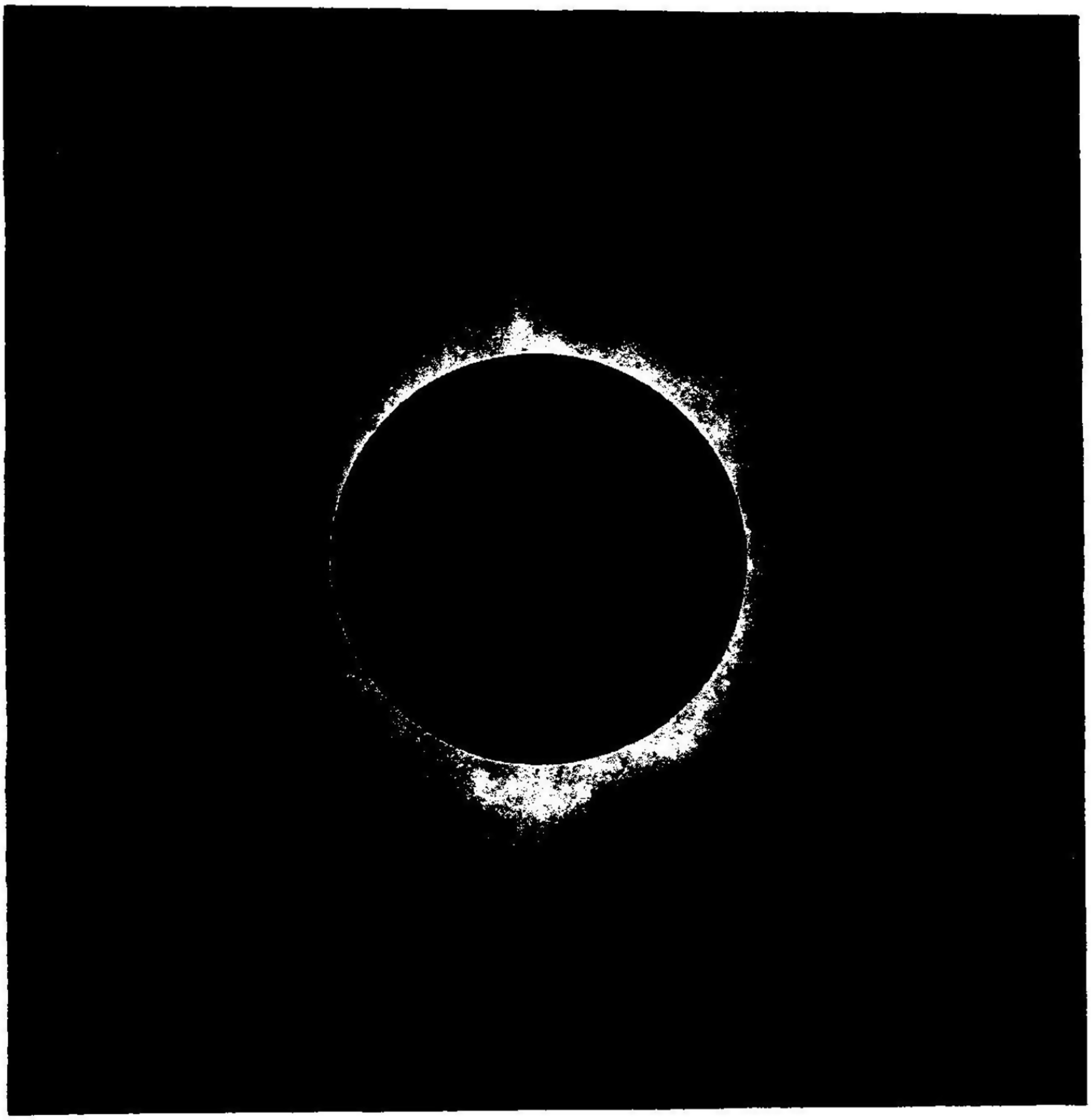


FIG. 41. Sun's corona as photographed during the solar eclipse of June 8, 1937. The duration of totality for this eclipse was the longest since 699 A.D. (National Geographic Society — U. S. Navy Eclipse Expedition: photographed by Dr. Irvine C. Gardner.)

In this particular solar eclipse the duration of the time of totality was the longest since 699 A.D.

The combination of circumstances as a result of which this phenomenon occurs takes place periodically. The interval between solar eclipses for a given location on the earth is 18 years, 10 days, 7 hours, and 43 minutes, a period of time that was called a *saros* by the ancient Babylonians.

If the lining up of the sun, moon, and earth takes place so that the earth is in the middle, an eclipse of the moon occurs, as shown in Fig.

42. Since the earth is much larger than the moon it casts a larger shadow, so that lunar eclipses occur more frequently and last longer than do eclipses of the sun.

Somewhat similar phenomena called transits and occultations occur with the other planets taking the part of the sun or moon. A transit takes place when Mercury or Venus comes between the earth and the sun. The appearance is simply that of a tiny black dot moving across the surface of the sun. An occultation of a planet or fixed star is said to occur when it is eclipsed or blotted out by the moon. In the course of time this happens to all the planets.

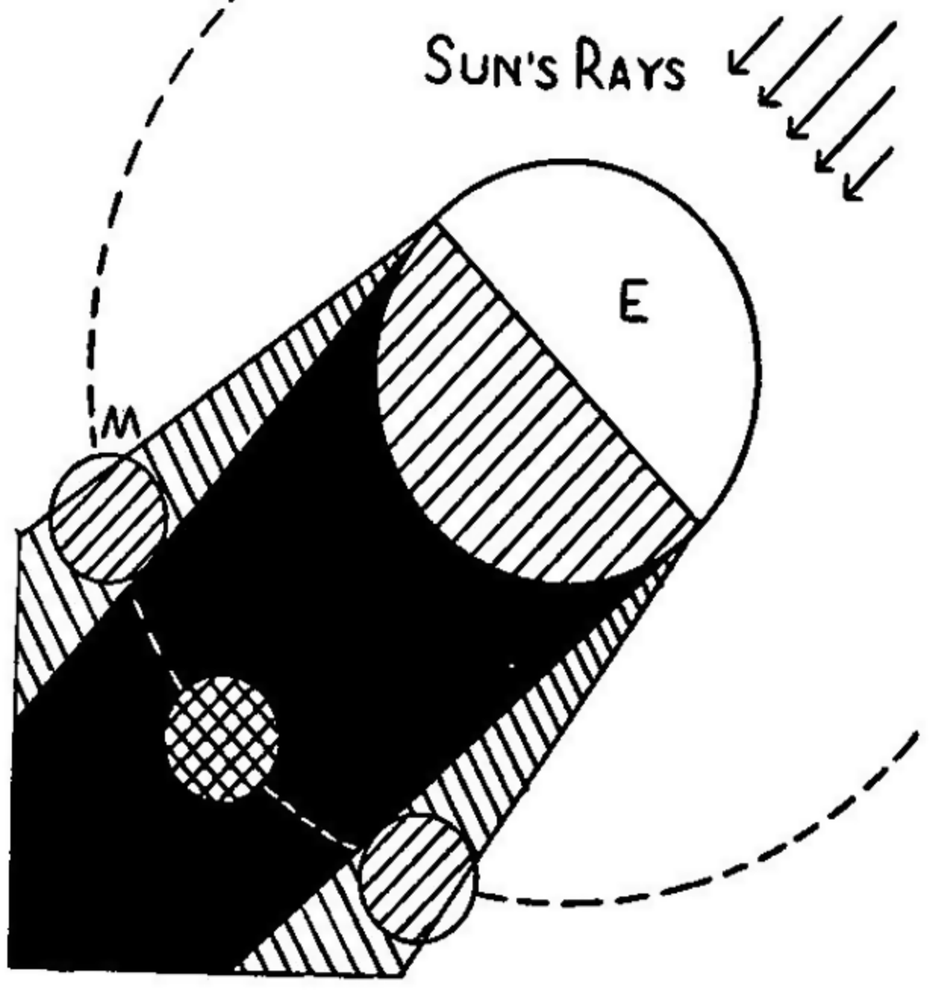


FIG. 42. Relative positions of sun, moon, and earth during a lunar eclipse.

Tides. — Finally, the moon represents the principal cause of tides on the earth. Consider, first, the earth and moon as stationary in the position shown in Fig. 43 (a). As a result of gravitational attraction ocean waters would tend to pile up to a height of several feet at points marked *H*. Since there is only so much water on the earth, this would produce low tides at *L* points. These places would tend to remain as they are with respect to the moon. Thus if the moon remained stationary, any given point on the earth's surface would pass through high tide every 12 hours, with times of low tide occurring halfway between. But the moon does not remain stationary. Since it revolves in the same direction that the earth rotates it will have moved to the dotted position shown in Fig. 43 (b) while the earth is making one rotation. This will make the times of high and low tide a little later than in the first case. Thus instead of high tide occurring exactly every 12 hours for a given location, it takes place about every 12 hours, 25 and a fraction minutes.

The sun also produces an appreciable effect on the tides although much less than that caused by the moon because of its greater distance from the earth. When the two bodies are located on either the

same side or opposite sides of the earth, their effects add together producing unusually high, or *spring* tides. On the other hand, if the sun and moon are in positions such that lines drawn from the two to the earth would be at right angles to each other, the gravitational attraction of one is working against that of the other and the resultant high tides are lower than normal. These are called *neap* tides. The height and occurrence of tides are also affected by local conditions of wind, atmospheric pressure, shore lines, and other factors. There is evidence to indicate that the tides act as a brake, slowing down the rotation of the earth about one-thousandth of a second per century. Thus it is gradually approaching the condition exemplified in our moon and in Mercury in which the period of rotation equals the period of revolution.

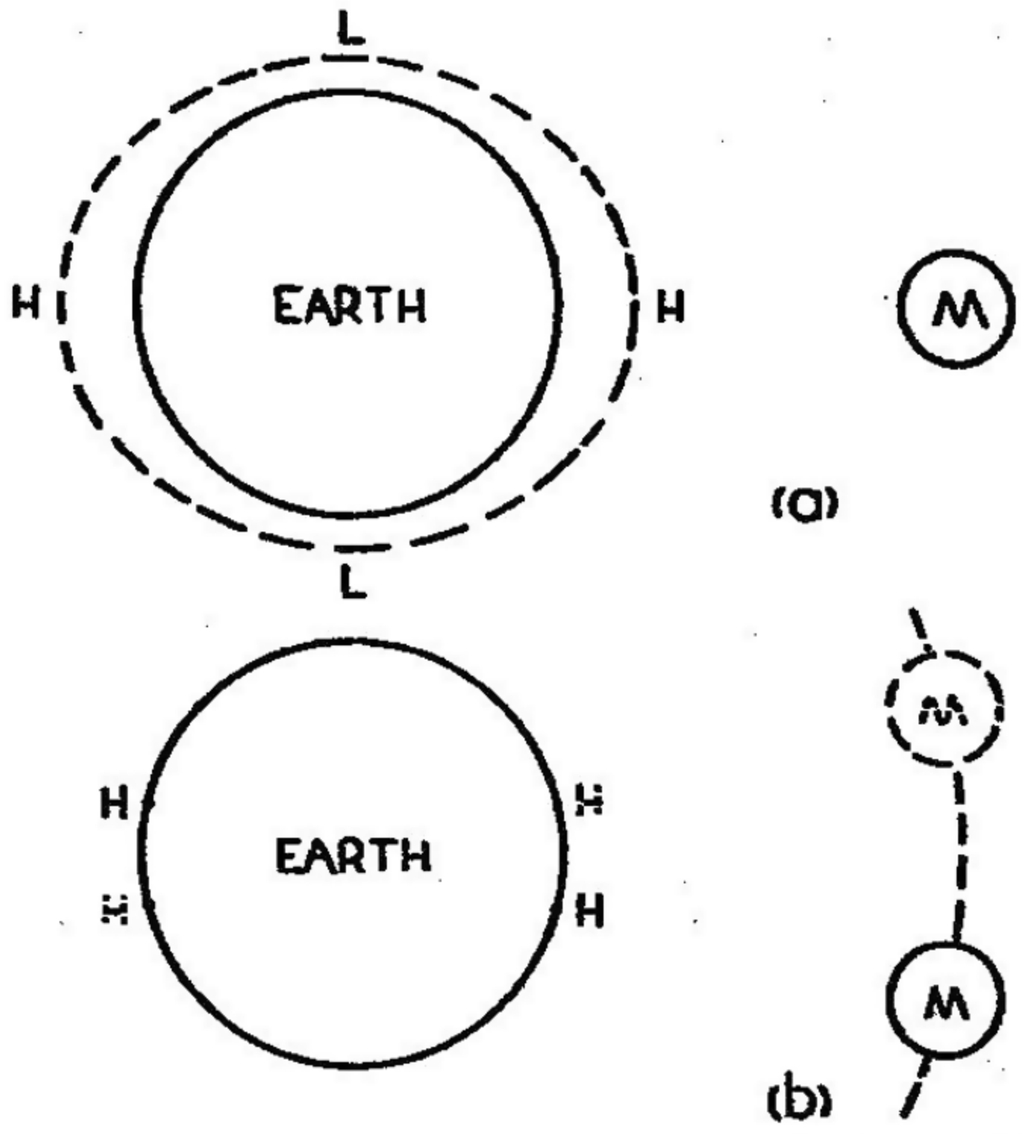


FIG. 43. Diagram illustrating production of tides on the earth.

QUESTIONS

Discussion

1. Discuss experimental evidence that you might use in convincing a skeptical individual of the truth of the following statements:
   a. The earth is roughly spherical in shape.
   b. The earth rotates about an axis.
   c. The earth revolves about the sun.
2. Describe in detail the astronomical method of determining the circumference of the earth.
3. Explain just how the following might be expected to affect the maximum range of temperatures existing on a planet:
   a. The composition and thickness of its atmosphere.
   b. The rate at which it rotates.
   c. The presence or absence of bodies of water on it.
4. Describe a Foucault pendulum and explain its use.
5. What changes in seasonal temperatures will occur in the following places during the next 130 centuries as a result of the earth's precessional motion?
   a. A point on the northern boundary of the United States.

b. A point on the equator.

c. A point in southern South America.

6. How would the seasonal variations in temperature of the places listed in Ques. 5 differ from their present values if the obliquity of the ecliptic were twice its present value? If it were 0 degrees?

7. Answer Ques. 6 substituting " variation in length of day and night " for " variations in temperature."

8. What effect on seasons will the earth's precessional motion have in 13,000 years? In 6,500 years?

9. Mention experimental evidence supporting or contradicting the following statements:

a. The moon has no atmosphere.

b. There is no water on the moon.

c. The moon is a star.

d. There is no life on the moon.

10. Explain the statement that the moon's period of rotation equals its period of revolution.

11. Suggest reasons for the common misconception that phases of the moon and eclipses have similar causes.

12. What phase of the earth would an observer on the moon be seeing when we see a new moon? Explain.

13. Compare the solar and lunar eclipses that we see with those that an observer on the moon would see with regard to (a) time of occurrence, (b) length, and (c) degree of totality.

14. Define and explain transits and occultations.

15. Why are successive high tides a little more than, rather than exactly, 12 hours apart?

16. Why is it that the moon rises about an hour earlier each night?

17. Under what sort of conditions would there be an eclipse of the sun every month? Explain.

18. Why are the following impossible?

a. An eclipse of a new moon.

b. A transit of Mars.

c. Solar eclipse at time of full moon.

d. Spring tides when the moon is half full.

e. Neap tides occurring simultaneously with an eclipse of the sun.

19. Draw a diagram to show the relative positions of the sun, the earth, and Venus when that planet appears to us as a thin crescent. As full. (See Fig. 7.)

Multiple Choice

1. If lines dropped vertically from two cities make an angle of 72 degrees when they meet at the center of the earth, the two cities are approximately — (a) 1,000, (b) 2,000, (c) 3,000, (d) 4,000, (e) 5,000 — miles apart.

2. The appearance in the heavens of a star inside the curve of a crescent moon, as on the flag of Turkey, (a) could only occur just before a solar eclipse, (b) could only occur at the time of equinox, (c) could only occur just before an eclipse of the moon, (d) could not occur at all.

3. On December 22, (a) a point on the equator has more than 12 hours of daylight, (b) the sun would appear directly overhead to an individual who lived on the Tropic of Cancer, (c) the South Pole would experience 24 hours of darkness, (d) the sun appears at its highest point in the sky to a resident of Argentina, (e) day and night are equal at the Arctic Circle.

4. Other things being equal, 26,000 years from now (a) the North Pole will point approximately at Polaris, (b) the South Pole will point at Alpha Centauri, (c) Vega will be the Pole Star, (d) the earth's axis will occupy a position in space perpendicular to its present position, (e) Arcturus will be the Pole Star.

5. An eclipse of the sun should occur every month at the time of a new moon if (a) the earth's axis were perpendicular to the plane of its orbit, (b) the plane of the moon's orbit were the same as that of the earth's orbit, (c) the moon traveled about the earth in a direction opposite to its present one, (d) the earth's orbit were circular instead of elliptical, (e) the orbits of the moon and earth were perpendicular to each other.

True–False

1. An observer located on the moon would see phases of "new earth," "half earth," "full earth."

2. The only planets for which transits can occur are Mercury and Venus.

3. If we could see it well enough we would expect that Pluto would exhibit the same series of phases as do the moon and Venus.

4. The sharp jagged appearance of lunar mountains suggests the absence of an atmosphere on the moon.

5. The Foucault pendulum is of value in demonstrating the rotation of the earth.

6. If the obliquity of the ecliptic were to be twice its present value, your home town would probably have warmer weather in June than it now experiences.

7. The southern hemisphere has its warmest weather when the earth is farthest from the sun.

8. Spring tides occur at a time when a line drawn from the sun to the earth would be perpendicular to one drawn from the moon to the earth.

9. An eclipse of the moon occurs when the shadow of the earth falls on the moon.

10. The fact that the craters on the moon are concentrated in the mountainous regions favors the hypothesis that they represent extinct volcanoes rather than the landing places of meteorites.

Chapter 7

THE EARTH AS AN ASTRONOMICAL BODY (Continued)

IN the preceding chapter our attention was focussed primarily on the relation of the earth to the other members of the solar system. In this chapter we shall concern ourselves more particularly with the composition and structure of the planet itself.

DIVISION OF THE EARTH INTO FOUR SPHERES

In discussing the makeup of the earth it is convenient to think of it as being divided into four parts or spheres. Two of these constitute the solid portion of the planet while the other two include the water and air respectively. We shall consider them in order, beginning with the interior of the earth.

The Centrosphere

Location. — The centrosphere may be defined as the interior of the earth, from the center out to a point some 40 to 60 miles beneath the surface. It is obvious that such a region cannot be investigated directly, so that what knowledge we have of it has been obtained by indirect procedures. However, as we shall point out from time to time in this book, indirect methods of acquiring information, when subject to the proper checks, may be quite as reliable as procedures which seem, on the face of them, less round about.

Properties. — The reader will recall that the nebular hypothesis of La Place pictured each of the planets as having begun its independent existence as a gaseous ring. This ring was then thought of as having formed into a more or less spherical mass which, as it cooled, first liquefied and then solidified to form the planet as we know it today. On this basis it was natural to think of the interior of the earth as still

being in a molten state, and so the term " crust " came to be applied to the solid outer layer which was thought of as floating upon a liquid interior. The evidence which led to the complete abandonment of the nebular hypothesis has already been discussed. Let us now look at some of the experimental facts which make it quite certain that the central portion of this planet possesses properties far different from those which we commonly associate with liquids.

Comparative densities of solid and molten rock throw some light on the problem. A cubic foot of the former is found to weigh more than a cubic foot of the latter. Thus, solid rock would not float on molten rock like ice on a pond, but would tend rather to sink to the bottom, that is, to the earth's center.

Also there is the matter of the frictional force that would be exerted by a liquid on the interior of a rapidly spinning sphere. It is a fact that a spherical metal shell containing a liquid cannot be made to spin for any length of time because of the frictional effect of the liquid on the inner surface of the ball. For this reason one might expect that the rotation of a molten-center earth would be slowing down much more rapidly than is indicated by the evidence.

A more quantitative kind of evidence comes from studies of the earth's density. In the previous chapter mention was made of procedures used in determining its size and mass. From these data the mass of an average cubic foot of the entire planet can be calculated. This average cubic foot is found to weigh about 5.5 times as much as a cubic foot of water. But when only materials near the earth's surface are used in the calculation, the average weight is only some 2.7 times that of water. This makes it evident that the substances which compose the interior of our planet must be considerably heavier than the average. Various calculations have placed the figure at from 10 to 12 times the density of water.

The study of earthquake tremors provides another avenue of approach to the problem. The procedures followed in these experiments will be described in the next chapter. It will be sufficient here to say that the results obtained are in agreement with those mentioned above, and indicate that the interior of the earth is more rigid than steel.

Examination of meteorites offers a suggestion as to the possible identity of the substances composing the earth's core. If one accepts

the idea that all the members of the solar system had a common origin, as implied in the planetesimal as well as the nebular hypothesis, then one might expect to find a similarity of structure among them. Meteorites, as man finds them, contain a far higher percentage of nickel and iron than do the upper layers of the earth's crust. But a meteorite is simply the core of what was once a much larger body, the outer portions having been consumed by the heat of friction as it traveled through the atmosphere. Thus it may very well be that the earth's core is also composed largely of a mixture of nickel and iron. Such a core would possess properties of density and rigidity in agreement with the experimental results previously discussed.

Another bit of supporting evidence for the nickel-iron picture should be pointed out. Iron and nickel are two of the three elements known to man which can be highly magnetized, and the observed magnetic properties of the earth are what one might expect if the central portion of the planet is made up of these two metals.

To summarize, in the light of modern experimental evidence, one may say quite conclusively that the interior of the earth is solid, has a high density, and is more rigid than steel. It also gives evidence of possessing the property of flowage, a characteristic which we commonly associate only with liquids. This is not contradictory to the rest of the picture, however, because under the enormous pressures which exist in the interior of the earth — as much as several million tons per square foot — materials which we know as solids might readily be expected to flow. The actual identity of the substances must not be stated too dogmatically, although we have described certain experimental facts that make it reasonable to think of the core as consisting of nickel and iron.

The Lithosphere

Location. — The lithosphere may be thought of as the outer shell of the solid earth some 40 to 60 miles in thickness. Since the deepest wells and mines of man do not reach more than two miles beneath the surface, here again most of our knowledge must be obtained by indirect procedures. Earthquake waves are of considerable value in this case also. Some important information is provided by volcanic materials which may come from considerable distances below the surface.

Even surface rocks are frequently a help in this connection since there is good geological evidence to indicate that many of these were at one time far beneath their present location.

Properties. — Investigations made by these and other methods show that about 35 of the known chemical elements are present in the lithosphere in amounts from .001 per cent up. Of these, the eight listed in Table 3 make up over 98 per cent of the total, with oxygen accounting for almost half.

TABLE 3. THE EIGHT MOST ABUNDANT ELEMENTS IN THE LITHOSPHERE

Element		*Percentage*
Oxygen	(O)	46.5
Silicon	(Si)	27.6
Aluminum	(Al)	8.1
Iron	(Fe)	5.1
Calcium	(Ca)	3.6
Sodium	(Na)	2.8
Potassium	(K)	2.5
Magnesium	(Mg)	2.1
		98.3

From man's standpoint one of the most important, as well as most impressive, characteristics of the surface of the lithosphere is its unevenness. The principal features of this irregularity are the ocean basins and the continental plateaus. Since, at the present time, there is more than enough water to fill the former, it encroaches to some extent on the continents, submerging the edges or so-called continental shelves. This overlapping is sufficient to make the total ocean surface area almost three times that of the dry land, although the actual ocean basins are only about twice as large as the continental areas.

Both the continents and the ocean basins are themselves marked by various relief features and topographical irregularities. The extremes of these are found in Mt. Everest, a peak of the Himalayas whose summit is 29,141 feet above sea level, and Tuscarora Deep off the coast of Japan where the ocean floor is 35,430 feet beneath the surface of the water. Thus the maximum range in altitude over the lithosphere is 64,571 feet or a little over 12 miles.

It should be noted that these relief features which appear so great

to us are actually quite insignificant when compared to the dimensions of the earth. In this connection someone has calculated that if a billiard ball were to be enlarged to the size of the earth, it would possess mountain ranges higher and valleys deeper than any now in existence on this planet.

It is primarily with the changes that have occurred and that are now taking place in the outer part of the lithosphere that we shall concern ourselves in Chapters 8 to 10 of this book.

The Hydrosphere

The hydrosphere, including oceans, streams, and the ground water which saturates a layer of earth just beneath the surface, forms an essentially continuous film about the lithosphere. The surface bodies of water cover almost ¾ of the approximately 197,000,000 square miles of the earth's exterior. They contain enough water, if distributed uniformly, to cover the entire planet to a depth of some two miles. As was mentioned above, at its greatest depth the ocean floor is farther below sea level than the top of the highest mountain is above it. In temperature, the oceans vary from 28° F. in the polar regions to 85° F. at the equator. Far beneath the surface, even in the lower latitudes, temperatures may be as low as 29° F. and remain almost constant throughout the year.

The tempering effect of the hydrosphere on earthly climates has already been mentioned. Because water both heats up and cools down more slowly than does land, oceans and lakes tend to keep temperatures within a narrower range than otherwise would prevail.

The Atmosphere

The fourth of the earth's parts is the gaseous ocean in whose depths we spend our lives. In discussing the atmosphere we shall concern ourselves with certain of its physical characteristics, with its separation into layers or strata, and finally with one of the most interesting phenomena it exhibits, the weather.

Physical Characteristics. — The thickness of the atmosphere cannot be stated exactly since it has no well-defined upper boundary but fades out somewhat gradually at high altitudes. However certain approximations can be given. It will be recalled that meteors cannot be

seen until they have encountered sufficient atmospheric friction to heat them to incandescence. Thus the fact that most of them become visible at altitudes of from 60 to 80 miles and a few as high as 125 miles indicates the existence of some air at these levels. Another source of information is the aurora borealis or northern lights, and its southern companion the aurora australis. These phenomena, which usually occur lower than 200 miles above the earth have occasionally been noted at heights between 200 and 600 miles, but they have never been seen at altitudes greater than 600 miles. Since they require the presence of air, the atmosphere above 600 miles would seem to be so thin as to be essentially non-existent.

The composition of the atmosphere by volume, near the surface of the earth, is given in Table 4. It should be noted that the various constituents are simply mixed together and do not combine chemically.

TABLE 4. PERCENTAGE COMPOSITION BY VOLUME OF THE ATMOSPHERE

Constituent	*Per Cent*
Nitrogen	78.03
Oxygen	20.99
Argon	.93
Carbon Dioxide	.03
Neon, Helium, Krypton, Xenon, Radon	.02
	100.00

In addition to the gases listed, varying amounts of water vapor, dust and other impurities are always present.

The fact that the atmosphere exerts pressure is another important characteristic. Since it is a gaseous substance, the air at any level is pushed down and compressed by all that is above. As a result of this we find that half of the total mass of the atmosphere is within 3½ miles of the earth's surface. Atmospheric pressure amounts to approximately 15 pounds per square inch at sea level and plays an important part in the determination of our weather.

This brief consideration of the atmosphere's physical properties will provide sufficient background for the material to be considered in the rest of this chapter. Some of these points will be taken up in greater detail in the portion of Chapter 15 which deals with air.

Stratification. — An important meteorological discovery of the

early part of the present century was the fact that definite layers or strata can be distinguished in the atmosphere. Five of these will be mentioned and the lower two considered in some detail.

Immediately surrounding the lithosphere and hydrosphere is the zone known as the *Troposphere*. The name itself signifies turning or turbulence and is descriptive of the constant agitation which this part of the atmosphere undergoes as a result of air currents. This continual mixing counteracts any tendency there might be for the constituent gases to separate out in order of their densities, so that the percentages given in the above table are essentially correct for the entire troposphere.

The upper boundary of the troposphere is defined in terms of a particular temperature phenomenon. If one were to rise vertically in a balloon from a point on the equator he would find that the temperature would decrease at the rate of about 16° F. per mile. This would continue for some 11 or 12 miles, at which point the thermometer reading would become about constant. Actually it would rise a little if he went still higher. The level at which this phenomenon of decrease in temperature with increase in altitude ceases is called the *Tropopause* and is the boundary between the troposphere and the next stratum. The troposphere is not the same thickness at all points. If the above balloon ascent had been made from one of the poles instead of from the equator, the tropopause would have been reached at an altitude of only four or five miles. Thus this lower atmospheric layer forms a kind of ellipsoidal envelop about the more solid portions of the earth. As a result of this varying thickness of the troposphere, the coldest natural earthly environment is not in or above the arctic regions, but directly over the equator at the tropopause. Here temperatures as low as —117° F. have been recorded.

Next above the troposphere is the zone so highly favored with publicity in recent years, the *Stratosphere*. As suggested above, the evidence indicates that temperatures throughout the 30 miles or so of vertical ascent that would take one from the tropopause to the top of the stratosphere, would rise a little. In this layer there are none of the winds, storms, and general turbulence that mark the one below. Consequently there is an appreciable tendency for the gases to separate out, with the heavier ones at the bottom and the lighter at the top.

The heavy gas, carbon dioxide, for example, seems to be entirely absent at altitudes above 15 miles. Hydrogen, the lightest of all gases, on the other hand is negligible in the troposphere but constitutes about 15 per cent of the air at the top of the stratosphere, and over 95 per cent from 70 miles up.

Just above the stratosphere is the *Region of Meteors.* It is in this

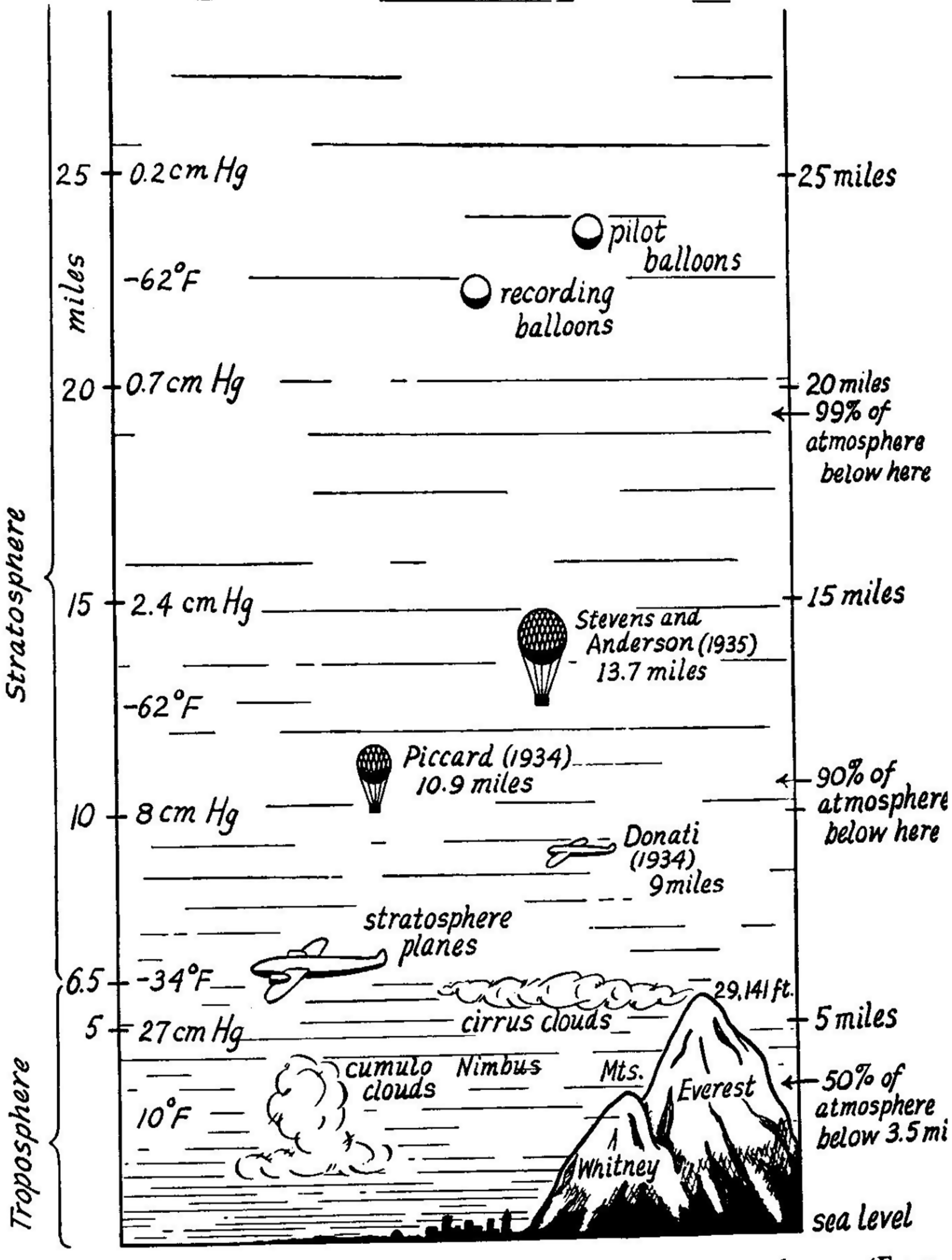


FIG. 44. Important facts about the troposphere and stratosphere. (From White's "Classical and Modern Physics," D. Van Nostrand Co., 1940.)

region that meteors in general encounter sufficient atmospheric friction to heat them to incandescence so that they become visible.

The next layer is known as the *Ionosphere*, or the *Kennely-Heaviside* layer. Here the atmosphere possesses certain electrical properties as a result of which radio waves of the frequencies employed in commercial broadcasting are reflected back down toward the earth.

The aurora borealis and aurora australis commonly occur at still greater altitudes and so the portion of the atmosphere above the ionosphere is known as the *Region of Auroral Displays*.

This completes the roll-call of the divisions of the atmosphere. Now let us return to the troposphere and consider briefly that commonest of all topics of conversation, the weather.

Weather. — The phenomenon that we call weather is confined to the troposphere and concerns the day to day variations in the condition of the atmosphere with respect to temperature, pressure, humidity, winds, and precipitation. We call it good weather or bad at any given time depending on whether or not we like the particular combination of these properties that is present. Although climate depends on the same set of characteristics, it concerns their long-time average rather than their daily fluctuations.

In considering causes of weather we will discuss briefly two general types of phenomena, those connected with atmospheric circulation or wind, and those resulting from the presence in the atmosphere of water vapor.

In the production of weather, as in so many other natural phenomena, the sun plays the leading role. Its part in producing winds results from the unequal manner in which various zones on the earth receive its heat. If we take the heating effect at the equator to be 100 per cent, it is but 90 per cent at the Tropic of Cancer, about 65 per cent along the northern boundary of the United States, and less than 40 per cent at the Arctic Circle. The same figures hold for corresponding locations in the southern hemisphere. Thus the air over the equator is heated to a higher temperature than that over other areas. Consequently it expands the most and rises. To take its place air must flow in toward the equator from north and south. If the earth did not rotate, there would be set up north winds near the equator in the northern hemisphere, and south winds in the southern, since

we name a wind for the direction from which it comes. But the rotation of the earth tends to deflect these currents and, in combination with the unequal solar heating and other factors, produces the general atmospheric circulation picture shown in Fig. 45. Around the earth at the equator is a belt of comparative calm called the doldrums. Here air currents are directed upward and so the region is one of low pressure. It is about 10 degrees of latitude or 700 miles in width. Adjacent to the doldrums are the trade-wind belts. Here the north and south currents mentioned above are deflected to form the northeast trade winds in the northern hemisphere and the southeast trades in the southern. Then at about latitude 30° N. and S., again there are belts of comparative calm. Here, in what are known as the horse latitudes, the air currents are directed downward, making them regions of high atmospheric pressure. The air that is flowing down in these belts is that which, after it rose over the equator, divided and flowed north and south until it was sufficiently cooled to sink again. The belts of westerly winds lie next to the horse latitudes. In these, as the name indicates, the prevailing winds are from the west. In the southern hemisphere they are very strong because there is so little land to break them up. The term " Roaring Forties " originated as a description of this belt between latitudes 40° S. and 50° S. Most of the United States lies in the westerly wind zone in the northern hemisphere. Around the poles are the so-called circumpolar whirls about which not a great deal is known. This rather simple general picture is greatly complicated in any given region by factors such as irregularities of the earth's surface, seasonal changes in temperature, and varying heat properties of land and water.

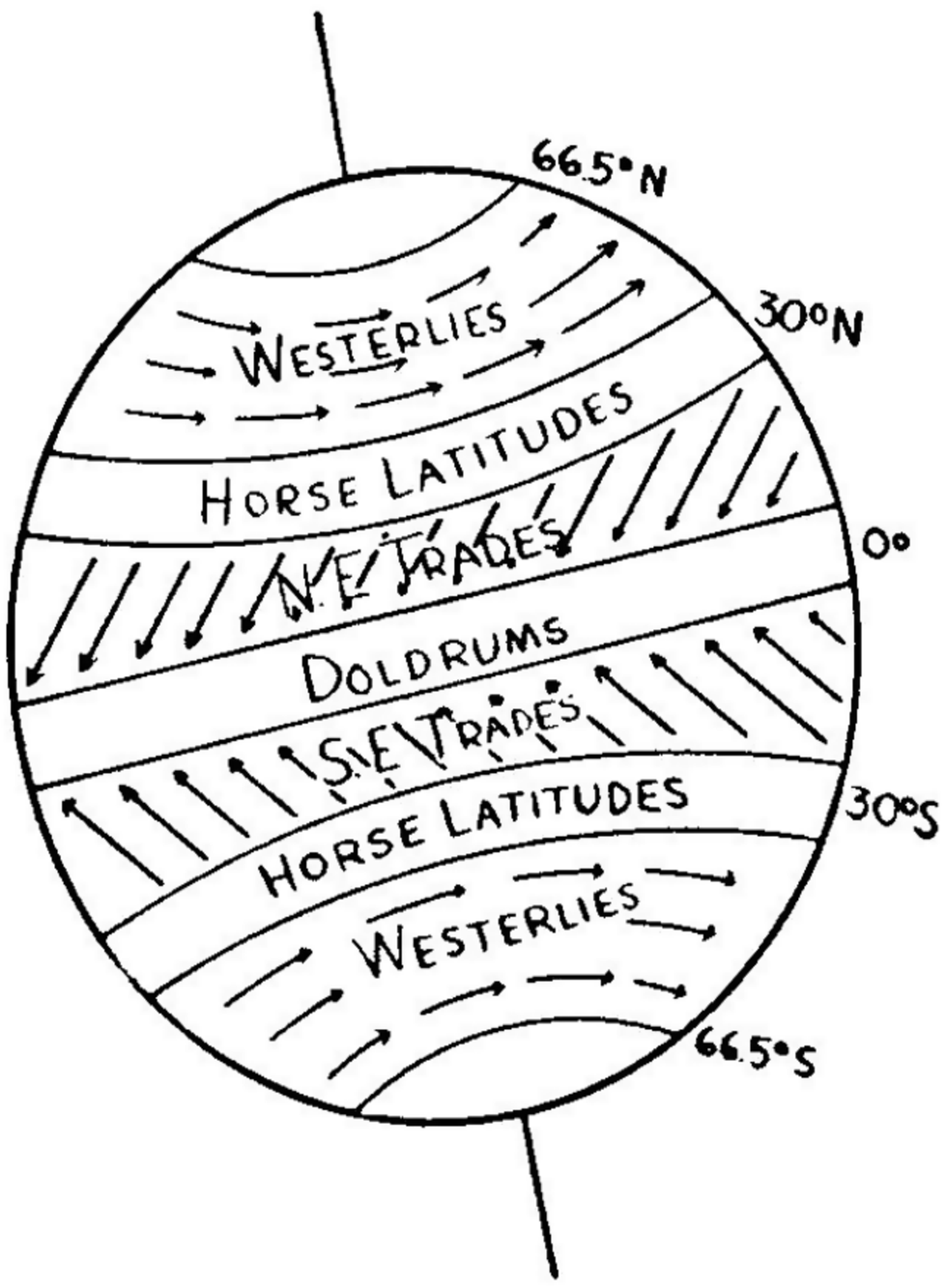


FIG. 45. The various wind belts on the earth's surface.

In considering the relationship of water vapor in the atmosphere to weather, two experimental facts may be stated as fundamental to the problem. These are,

1. Water evaporates into the atmosphere;
2. The maximum amount of water vapor that can be present in any given volume of air increases with the temperature of the air.

Now keeping these facts in mind let us assume that a certain portion of the atmosphere is at 70° F. and is saturated, that is, it contains the maximum amount of moisture at that temperature. Then we will suppose that the temperature drops rather quickly to 65° F. Since at the new temperature the air cannot hold as much moisture as it could at 70° F., the excess condenses out, making itself known to us in the form of some sort of precipitation.

Before describing the various forms that this condensed moisture can take, we must define two terms. One of these is *dew point*, which is the lowest temperature to which a given volume of air can be lowered without condensation occurring. The other, *relative humidity*, expresses the amount of water vapor in the air at any given time in comparison to the quantity it can hold at that temperature. Thus a relative humidity of 75 per cent means that the atmosphere contains 75 per cent as much moisture as it would contain if it were saturated at the same temperature.

In the above example we mentioned a special case of precipitation forming when air was cooled from 70° to 65° F. Generalizing, we may say that for any kind of precipitation whatever to form, it is necessary that air be lowered in temperature below the point at which it would be saturated, that is, below dew point. It will be clear that when the temperature reaches the dew point, the relative humidity equals 100 per cent. If the condensation occurs rather rapidly at a temperature above freezing and the moisture forms into drops, we ordinarily become aware of it in the form of *rain*. Such drops form about dust or other particles as nuclei. If the moisture condenses in the form of very small particles they may remain suspended in the atmosphere as a *cloud*. Very high clouds consist of ice particles. If rain, as it falls, passes through an atmospheric layer that is below freezing, it will arrive at the earth's surface in the form of *sleet*.

When a cloud, as pictured above, forms near the ground instead of high in the air we give it a different name and call it *fog*. When raindrops are forced, by air currents, up and down several times between the zone where they formed and a higher one in which freezing temperatures exist, successive layers of ice accumulate on the particles and eventually they fall as *hail*. If the condensation occurs in a region of freezing temperature, *snow* forms instead of rain. This leaves *dew* and *frost* which result from condensation occurring close to the ground, the former when the air is above freezing and the latter when it is below.

Various combinations of circumstances may cause the temperature changes which lead to the formation of precipitation. Two examples will be mentioned. In the region of the doldrums, described above, air is moving upward and so is being lowered in temperature. Also, since most of this area is over the ocean, this rising air has a high moisture content. Consequently one would expect this to be, on the average, a region of considerable rainfall, as it is. The horse latitudes, on the other hand, might be expected to experience much less rainfall since the prevailing air currents are vertically downward. Thus the air is moving to levels of successively higher rather than lower temperatures.

When the ingredients of presence or absence of sunshine, and daily temperatures, are added to those just discussed and the whole well mixed together, there emerges the commodity which we call weather.

Weather Forecasting. — Notwithstanding Mark Twain's famous statement to the contrary, man has done in the past and is doing today a great deal about the weather. Some of his efforts in this connection have been directed toward attempts to control weather, but these have met with little or no success, except within very small limits. Air conditioning, for instance, might be thought of as weather control on a small scale.

Much more of man's energy has been devoted to the problem of foretelling weather, and with much greater success. The origin of his attempts to predict weather is lost in antiquity. Even the recorded history of his investigations along this line is long and crowded, dating from about the seventh century B.C. Since that time the bases upon

which weather forecasts have been made have included everything from the arrangement of the stars and the habits of insects, to the groundhog's shadow and man's own rheumatic twinges. The history of the growth and development of this great mass of weather lore is a fascinating story but one which space does not permit us to include here.

Modern predictions are based on accurate, carefully collected data. Since the beginning of organized forecasting, the fundamental principle of prediction has been the observed fact that *weather travels.* In the United States the general direction of travel is from west to east. The weather man's first step in preparing a forecast is to get as complete a picture as possible of existing weather conditions over as large an area as possible. To supply this picture the United States Weather Bureau maintains its vast organization. As a typical example we shall consider the weather station at Cleveland, Ohio.[1] This station, like a large majority of those from which reports are obtained, is located at an airport. The meteorologists at the Cleveland bureau receive data from 595 weather stations scattered over the United States. Almost all of these points report either every hour or every six hours with a few sending in their information at irregular intervals. Other reports come from Canada and Mexico, and from ships in the Atlantic and Pacific oceans and on the Great Lakes. The data received include temperature, atmospheric pressure, humidity, wind and cloud conditions, ceiling, visibility, and precipitation. From these extensive reports a complete weather map is made every six hours and auxiliary charts at the intervening three-hour periods. A daily printed map, based on the 1:30 A.M. map is made and placed in the mail at 5:30 A.M.

On the weather maps, two features are outstanding. They are shown diagrammatically in Fig. 46. The one labeled " low " is a kind of huge whirlpool, known in meteorolgy as a *cyclone.* Atmospheric pressure is low at the center of a cyclone and increases toward the outer edge. This is due to the fact that the air is moving upward in a spiral fashion. Since it is rising, this air is being cooled and so any moisture it contains is likely to condense out as some form of precipitation. Thus the kind of weather that we call " bad " usually accompanies

[1] Information on Cleveland Weather Bureau through courtesy of C. G. Andrus, senior meteorologist, July 23, 1941.

such a low-pressure area, its approach being heralded by a " falling barometer." An additional reason why approaching precipitation is preceded by a decrease in atmospheric pressure is the fact that air which contains water vapor has a lower density than air which has no water vapor in it.

The other feature is essentially the opposite of the one just described. It is called a " high " or anticyclone. At its center atmospheric pressure is high because the air is moving spirally downward

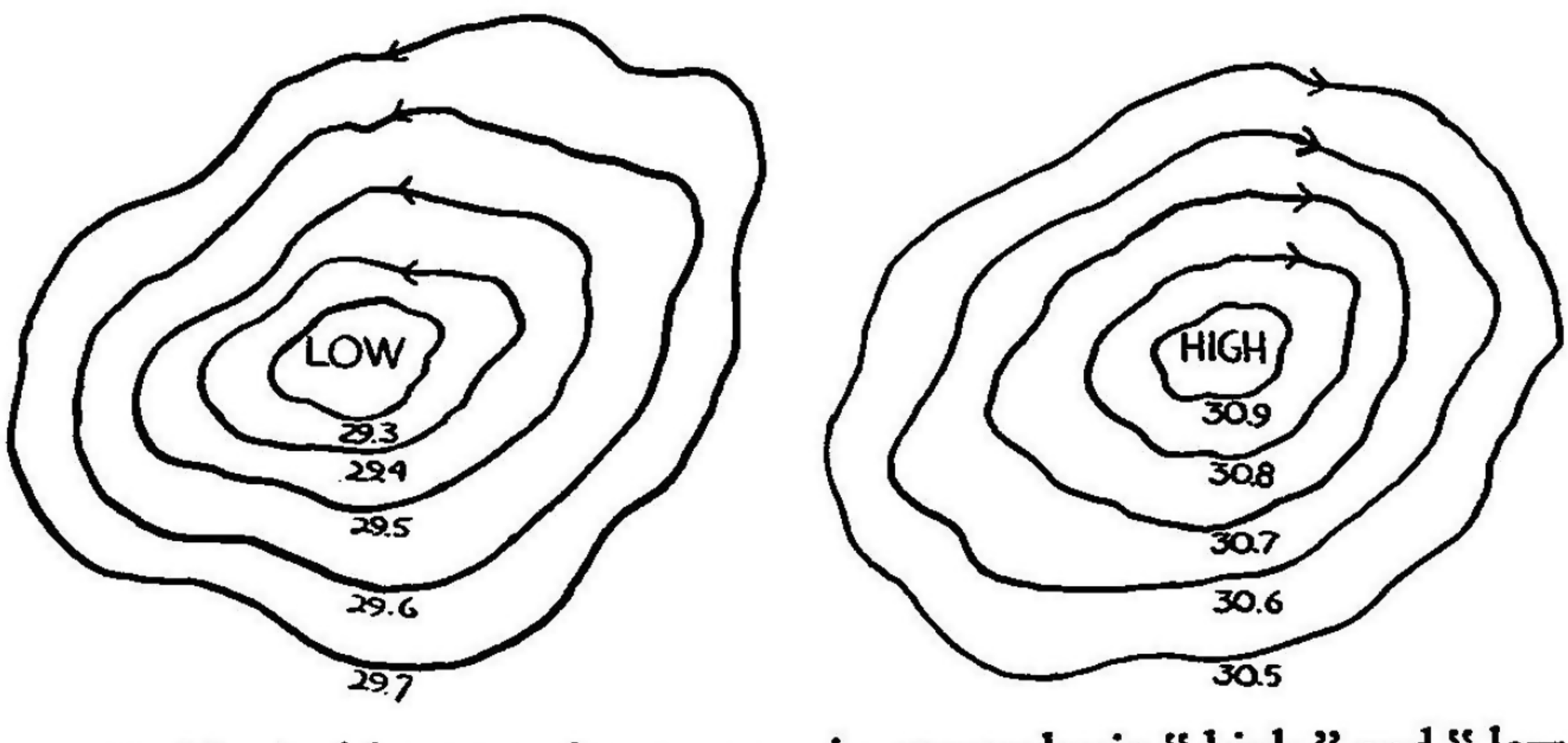


FIG. 46. Typical barometric pressures in atmospheric " high " and " low."

and so is encountering increasing temperatures. Precipitation is less and less likely to occur, and a " rising barometer " in general indicates that weather of the sort we call " good " is approaching. A typical weather map is shown in Fig. 47. Note the location of " highs " and " lows," particularly the region of low pressure in the Great Lakes area. This marked the center of a very heavy rain and snow storm.

It will be noted that in the procedure as we have discussed it thus far, the forecaster makes no attempt to discover any of the physical causes behind the weather. He merely observes present conditions, compares these with past experience, and proceeds to predict the immediate future. His technique may be illustrated in the following manner. Consider the case of a man who lives on the shores of a river whose waters are polluted with waste materials from a factory upstream, the amount and character of the impurities varying from day to day. Also suppose that it is desirable that he be able to predict at any time just what the extent and kind of pollution will be a certain length of time ahead. It is conceivable that by frequent testing of

the water, by keeping accurate records and charts, and by studying trends for a period of several years, he might learn to make such predictions with considerable accuracy and still know little or nothing about the actual causes of the pollution. On the other hand, if he should visit the factory and learn the causes, he might increase the accuracy of his forecasts.

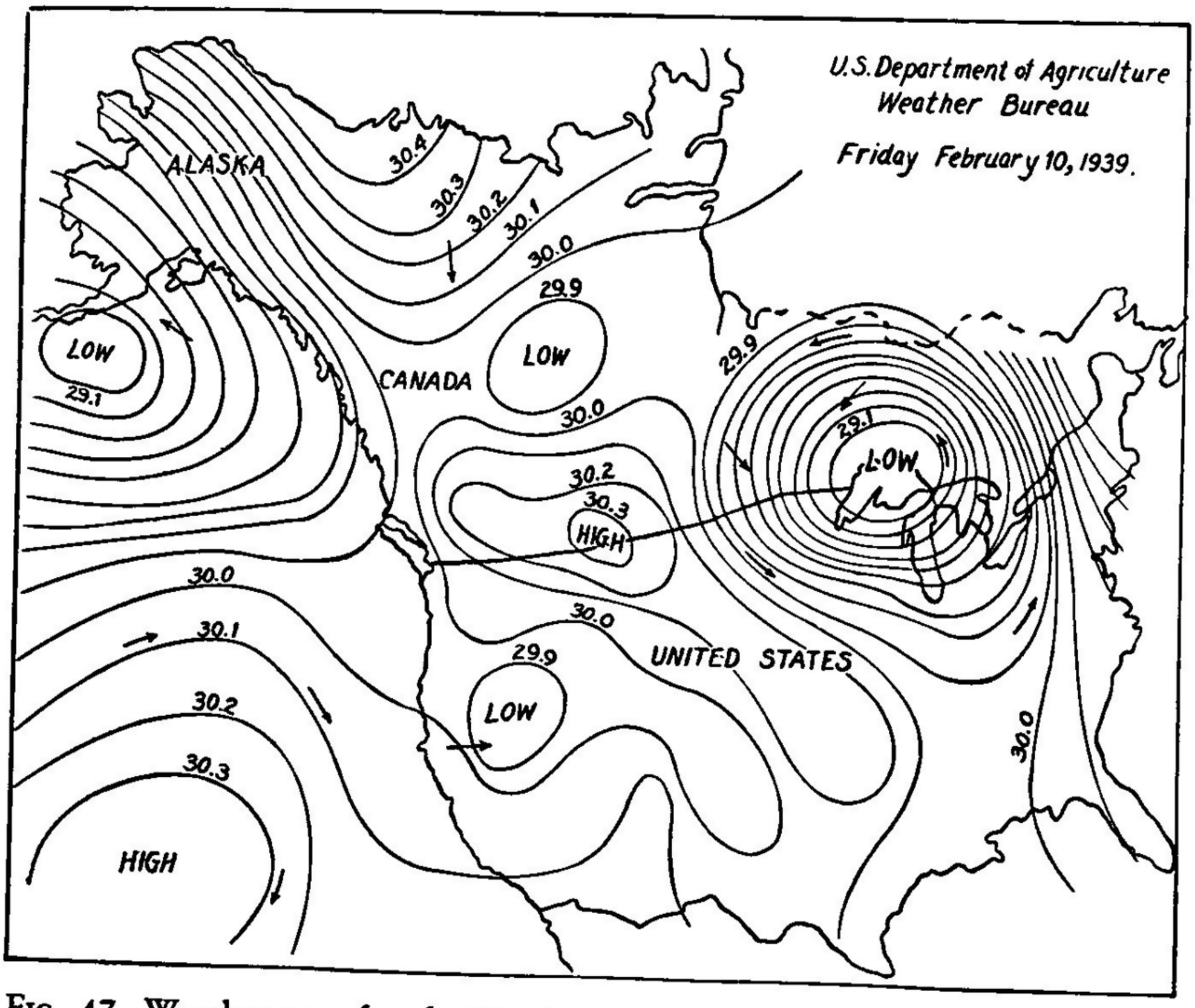


FIG. 47. Weather map for the North American continent for Friday, February 10, 1939. (From White's "Classical and Modern Physics," D. Van Nostrand Co., 1939.)

The first procedure is analogous to that employed exclusively in weather forecasting until comparatively recent times. Within the last few years techniques have come into use which do make an attempt to learn what causes weather. The method originated in Norway and is commonly called "Air-Mass Analysis." It does not replace the procedures just discussed but supplements them. Air-mass analysis treats the atmosphere as a sort of huge three-dimensional battleground in which a number of more or less separate and distinct masses or "chunks" of air are at war. These gigantic foes are in conflict because of their different properties, principally of temperature and

humidity. Now at first thought it may seem quite impossible to have distinct masses with fairly well-defined boundaries in anything as fluid and continuous as the atmosphere. That it can be so considered may be illustrated in the following manner. If a quart of colored water is suddenly dashed into a large tub of clear water, the former will behave for a little time very much like a solid projectile. That is, it will penetrate as a whole, maintaining briefly its general shape. Then as its speed lessens, little swirls and eddies will develop along the bounding surface between the colored water and the clear.

In the case of weather formation one of these masses of air begins its march after having been stationary over a particular region for some time. While thus located it has acquired the general temperature and humidity characteristics of that area. Air which remains over the Gulf of Mexico for some days or weeks has a comparatively high temperature and large moisture content, and such a portion of the atmosphere is known to meteorologists as Tropical Gulf air. On the other hand, atmosphere over northern Canada, called Polar Continental air, tends to be cold and dry. In like fashion Polar Atlantic air and Polar Pacific air are both moist but at a low temperature. Then when these moving masses of air, having different characteristics, come into contact atmospheric disturbances develop along the bounding surfaces, called fronts, just as the swirls and eddies developed between the colored and clear water masses. These disturbances constitute the regions of rapidly changing conditions as a result of which we have weather changes.

Since these fronts extend to considerable heights, their study is a three-dimensional problem and cannot be carried on by making observations only on the surface of the earth, so that techniques for the collection of meteorological data at high altitudes have been increasingly necessary of recent years. Even before the wide application of air-mass analysis methods to weather forecasting, considerable information had been gained about the atmosphere at varying distances above the earth. Some of the early investigations were carried on with kites, some with captive balloons, and some with man-carrying balloons. By far the most important device for this work has been the sounding or meteorological balloon. This consists of a four- or five-foot gas bag carrying a recording instrument. The record, which is scratched automatically on a copper tape, includes temperature, humidity, and at-

mospheric pressure, at definite intervals throughout the time the balloon is in the air. The outfit ascends until the bag bursts after which the recording device descends by means of a parachute. With these balloons routine data is collected up to altitudes of from 12 to 16 miles. The world's record for such an ascent is 22 miles. The information gained in this fashion is immensely important from the standpoint of increasing our knowledge about the upper atmosphere but not so valuable as far as weather forcasting is concerned, since the instruments may not be found and returned to the weather station for days or even weeks after the flight.

The advent of air-mass analysis made necessary the use of some more rapid procedure. For several years the data used in forecasts were collected by airplane flights. These were made daily from some 36 airports in different parts of the country. This method had two disadvantages. In the first place the coverage of the nation was scarcely adequate, and in the second place, bad weather frequently interfered with the flights. Thus the very days that the data would have been most valuable were the days the planes were unable to collect it. More recently there has been developed a new type of sounding balloon which has now replaced the airplane flights. Instead of recording temperature, humidity, and atmospheric pressure, the instruments carried in this outfit automatically broadcast these data at 1-minute intervals. Thus the weather man has the information for immediate use in preparing his forecast. Here again the instruments descend by parachute after the gas bag bursts. Red parachutes are used to make the devices easier to find and a reward is offered for their return to the weather bureau.

The weather man of today has at his disposal an enormous amount of information about the atmosphere, both on the earth's surface and far above it. In spite of the things we are prone to say about him, his forecasts have been increasingly accurate over recent years and can be counted upon today to be right some 90 per cent or more of the time.

Questions

Discussion

1. Define and locate the centrosphere, lithosphere, hydrosphere, and atmosphere.

2. State and discuss reasons for abandoning the belief in a molten interior of the earth.

3. What experimental evidence is there that the center of the earth is ten to twelve times as dense as water?

4. On what sort of evidence is the suggestion based that the center of the earth is composed of nickel and iron?

5. State the four most abundant elements in the lithosphere, and the approximate percentage of each.

6. What are the continental shelves?

7. Discuss the oceans with regard to extremes of depth and temperature.

8. Explain the tempering effect on climate of large bodies of water.

9. What evidence is there of the presence of atmosphere at heights of more than 125 miles above the earth?

10. List constituents that make up 99.9 per cent or more of the atmosphere.

11. Name the various strata in the atmosphere and state the physical characteristics of each.

12. Discuss the particular differences between the troposphere and stratosphere.

13. Why do gases separate out in order of their density in the stratosphere and not in the troposphere.

14. If you were making a balloon ascent what instrument would you need in order to know when you reached the top of the troposphere? Explain.

15. Explain the fact that hydrogen makes up a high percentage of the upper part of the stratosphere when it is almost entirely absent from the troposphere.

16. What determines the boundaries of the region of meteors?

17. Discuss the role played by the sun in the formation of weather.

18. Show by a diagram the location of the various wind belts on the earth's surface.

19. Explain the formation of rain, hail, snow, sleet, fog, frost, and dew.

20. Find out what meteorological data are received at your local weather bureau and by what methods.

Multiple Choice

1. All of the following have been suggested with respect to the centrosphere. Which *one* is *most* conclusively substantiated by modern evidence? (a) The centrosphere is a liquid upon which the lithosphere floats. (b) The centrosphere is a very hot gas. (c) The centrosphere is an alloy of nickel and iron. (d) The centrosphere has a density ten to twelve times that of water.

2. The greatest depth of the ocean floor below sea level (a) is about half the height of the highest mountain, (b) is about 10 miles, (c) is about half as great as the distance to the top of the stratosphere, (d) is a little greater than the top of the highest mountain is above sea level, (e) is about twice the height of the highest mountain.

3. If the temperature and relative humidity are 80° F. and 99 percent

respectively, (a) it is raining, (b) a rise of 10 degrees in the temperature will cause rain, (c) the temperature must drop before it can rain if no more moisture is absorbed, (d) any rise in temperature will necessarily be accompanied by a rise in the relative humidity, (e) the relative humidity must decrease before rain can fall.

4. If it were not for the rotation of the earth, the belt of northeast trade winds would be a belt of — (a) north, (b) east, (c) south, (d) west — winds.

5. If a barometer and a thermometer were to fall out of a balloon at a height of 7 miles over the equator — (a) the readings of both would increase, (b) the readings of both would decrease, (c) the barometer reading would rise and the thermometer fall, (d) the thermometer reading would rise and the barometer fall — all the way to the earth.

True–False

1. Data obtained from earthquake waves tend to support La Place's idea of a molten central portion of the earth.

2. The polar regions are colder than the tropics because of the greater thickness of the troposphere over the former which keeps out much of the sun's heat.

3. The upper boundary of the stratosphere is called the tropopause.

4. Weather is confined to the troposphere.

5. The Region of Meteors may be defined as the zone in which these masses of metal and rock are believed to originate.

6. The gases of the atmosphere show a greater tendency to separate out in layers in order of their densities, in the stratosphere than in the troposphere.

7. The amount of atmosphere that surrounds a planet is determined in part by the mass of the planet itself.

8. Frost and dew differ in that the former falls while the latter is formed near the ground.

9. The kind of weather that we call good is usually associated with a downward flow of air currents.

10. Since water heats up and cools down more slowly than does land, along the seashore breezes from the ocean toward the land would be expected to be more common during the morning than in the late afternoon.

Chapter 8

FACTORS WHICH CHANGE THE SURFACE OF THE EARTH

IN the preceding two chapters we were concerned with the earth as it now is. With this one we begin a survey of the events which led to its present state. In other words, we are about to ask the earth "how it got this way." In doing this we shall first describe the various types of rock which compose the earth's crust, second consider the agents which are at work changing its surface, third examine the clues to earth history, and finally see something of what that history actually has been. This particular chapter is concerned with the first two of these points.

BEGINNING OF EARTH SCIENCE

Geology, the physical science that deals with these matters, is one of the youngest members of the natural science family, although the phenomena with which it is concerned have been in operation since the beginning of earth time. Some branches of science had attained a considerable growth by the time of the Renaissance. Astronomy, as we have already seen, was one such field. Mathematics was another. The field of geology, on the other hand, was almost non-existent as little as 125 or 150 years ago. Actually this is not particularly surprising since as recently as 75 years ago the belief was still widespread that the earth had been created within the last 6,000 years in exactly its present form, that is, with all rock layers, fossils, mineral deposits, and so forth quite intact. So naturally there was little reason for the existence of a science which would concern itself with earth history when one assumed to begin with that the planet had almost no history. Geology began to come into its own when man finally recognized that Mother Earth would have to admit to an age far greater than the few

thousand years with which he had been so gallantly crediting her. In fact the best evidence today indicates that her birthdays must be numbered in billions rather than in thousands. This evidence will be discussed in the next chapter.

Types of Rock in the Earth's Crust

The various chemical elements which are found in the earth's crust were discussed in the preceding chapter. It will be recalled that eight are found to compose over 98 per cent of the lithosphere, with oxygen alone accounting for almost half. Another basis of classification of these materials groups them into four types of rock, namely igneous, sedimentary, metamorphic, and mantle.

Igneous Rock

The very name of igneous rock implies a connection with fire and heat. Actually, it is simply rock that was once molten and has now cooled and solidified. The source of the heat which produces this molten rock, called magma, is not completely understood. At one time, as we have seen, it was believed that the entire central portion of the earth was molten, that is, was simply the fraction that remained of a globe that was once entirely in the form of a hot fluid. Another hypothesis attempted to explain the heat as being the result of contraction of the earth. Calculations show that neither of these is at all adequate as far as providing enough heat is concerned. The theory that at the present time seems to be the most satisfactory, or as someone has said, the least objectionable, pictures the heat as resulting from the energy emitted by radioactive substances like uranium and radium. Figures based on estimated amounts of these materials present in the earth's crust indicate that sufficient heat would be provided by this phenomenon, if too much of it did not escape from the surface.

Since about 95 per cent of the earth's crust is composed of igneous rock, its average composition is essentially the same as that already mentioned for the lithosphere as a whole. Granite and basalt are two of the more familiar forms of this kind of rock. The former composes large fractions of young mountain ranges like the Rockies. Basalt is darker in color and is the heaviest of the rocks that are common near the surface. Layers of it underlie the oceans and compose

most oceanic islands. It will become evident as we proceed that igneous rock is the fundamental type and serves as a primary source of the others.

Sedimentary Rock

This type of rock, as the name suggests, is the product of settling of some sort. It consists mostly of layers of material which at some time or other settled to the bottom of bodies of water, usually oceans. Although these sediments compose less than five per cent of the earth's crust, they are spread out over some three-fourths of the total land area. Thicknesses vary from zero to eight or ten miles. The materials which make up the oldest of the sedimentary rocks came exclusively from igneous rock which was broken up by one means or another. Igneous material forms the primary source of even later sedimentary formations although these also contain other types of rock. From the standpoint of geological history, the principal importance of sedimentary rock lies in the fact that it contains the fossils which are the essential clues to these past events.

Metamorphic Rock

The earth's crust contains other rock deposits whose structural features and minerals differ from either of those discussed above. Investigation shows them to be rocks which once were igneous or sedimentary but which were changed into their present form. Consequently, they are called metamorphic rocks. The alterations which occur are both physical and chemical in nature. The principal agent of physical change is pressure which may squeeze the existing layers into new forms and shapes. Chemical agents include water and various gases which by their action produce new materials. Slate and marble are examples of metamorphic rock. Slate comes from an original deposit of mud which formed shale and then was changed into its present state. Marble is metamorphosed limestone.

It should be noted that any given bit of rock material may go through a cycle which would include all three of these types. Metamorphic rock, for example, may be forced down to a zone where temperatures are high enough to melt it. Thus it becomes magma which later solidifies and becomes igneous rock. This in turn may become

exposed, be broken up and deposited by water as a sediment. Later it may be acted upon by the agents mentioned and become metamorphic rock again.

Mantle Rock

Much of the material which we see about us on the earth's surface cannot be included in any one of the above three forms. There remains the mantle of sand, mud, pebbles, and organic substances, whose upper foot or two we commonly call soil. This mixture is named mantle rock. Its thickness varies from zero in places where bedrock is exposed to hundreds of feet in jungle regions where decay proceeds at a rapid rate.

Having defined the rock types which are found in the upper crust of the earth, let us now turn our attention to the main business of the chapter, as indicated by its title. Although poets sing of the "everlasting hills," and the Rock of Gibraltar is pictured as the very acme of solidarity, earth structures are anything but permanent. There are two groups of agents which are now and always have been busily engaged in altering the face of the earth in all sorts of ways. One of these comprises a set of what we may call "relief-levelling" factors. These, if unhindered, would eventually reduce the high places and fill up the low areas of the earth's surface until it was entirely covered with water. Opposing the gradation factors is a group of "relief-building" agents whose tendency is just the opposite. Both sets are always at work although, as we shall see, at any given time one or the other may be dominant.

Relief-Levelling Factors

The total work of these agents is usually included in the term "gradation." It includes the piecemeal removal and wearing down of high regions, and the accumulation and deposit of the materials in low places.

Temperature Variations

An appreciable amount of rock erosion occurs because of changes in temperature. If you were to pour boiling water into a cold thick glass tumbler, you would not be greatly surprised to see the glass

break. Two experimental facts explain this result. Fact number one is that glass, like most substances, expands as its temperature rises. Fact number two is that glass is a poor conductor of heat, that is, heat does not travel through it as readily as it does through materials like iron and copper. When the inside of the glass comes into contact with the hot water, it expands rather suddenly. Since the heat is not transmitted rapidly, the outside expands very slowly. As a result strains develop and the glass breaks.

Similarly, rock is a poor conductor of heat. In localities where temperatures vary considerably between noon and midnight, outer layers of rock alternately expand and contract and finally crack and peel off, exposing new surfaces to the same forces. In some places these variations are as much as 100° F. within a 24-hour period. As a result large boulders are gradually worn away and disappear. Massive outcroppings of rock are eroded and the material piled up at the base where it may be carried away by other weathering factors. Even when the large rock masses are not actually broken up, temperature variations open innumerable tiny cracks which admit water, carbon dioxide, and oxygen, and so aid the erosional work of the other factors.

Water

There are several different ways in which water enters into the total gradational work of the world. When it freezes, for example, it expands to about eleven-tenths of its former volume. If it happens to be tightly confined when this occurs, pressure as high as a ton per square inch may be exerted. Everyone is familiar with the unfortunate results of such expansion in water pipes and automobile radiators. Similarly, when it freezes in rock crevices, fragments may be broken off. In fact, practical use is made of the phenomenon in certain European stone quarries. During the day holes are drilled in the rock along the desired lines of cleavage and filled with water. At night the water freezes, expands, and quarries out the huge blocks.

The most effective of all the erosional agents is running water. Its source is rainfall or snowfall, of which the total annual volume is estimated to be some 30,000 cubic miles, sufficient to cover the entire earth to a depth of about one foot. Just now we are interested in the portion which runs off comparatively rapidly. This fraction va-

ries greatly from one place to another depending on such factors as slope, rate of rainfall, evaporation, condition of surface, nature and amount of vegetation, but on the whole averages about a third. As this water makes its way down slopes to form into ever larger and larger streams, it carries with it considerable amounts of various earth materials. These are removed from one place, transported for varying distances grinding and abrading each other as they go, and are



FIG. 48. The formation of gullies, the beginning of stream erosion. (Photo by W. O. Thompson.)

finally deposited at a lower point, thus aiding in the general levelling process. The importance of this transfer is evident from the fact that the Mississippi River alone carries out and deposits in the Gulf of Mexico over 400 million tons of mud and silt every year.

Previous to a time a little over a century ago it was generally assumed that a valley was the result of some violent action such as an earthquake, and then was occupied by a river later on. The principal reason for this belief was that it " just wasn't common sense " to think that a river would be able to carve out the valley or canyon through which it traveled on its way to the ocean. But gradually, thoughtful observers began to notice a number of striking features about river valleys. Some of these were, that in general a rather definite relationship exists between the size of a river and its valley, that streams almost always join together at the same level, and that smaller streams invariably flow into larger ones at angles which are acute with the upstream direction. Cases were actually observed in which rivers broke from their channels, flowed across valleyless regions, and

very shortly cut new channels which were big enough to contain them. In the light of these experimental facts and others, it was finally recognized that rivers actually do carve out their own valleys. And once again, something that wasn't common sense turned out to be true. Now let us trace briefly the sequence of events in the formation of a typical river valley.

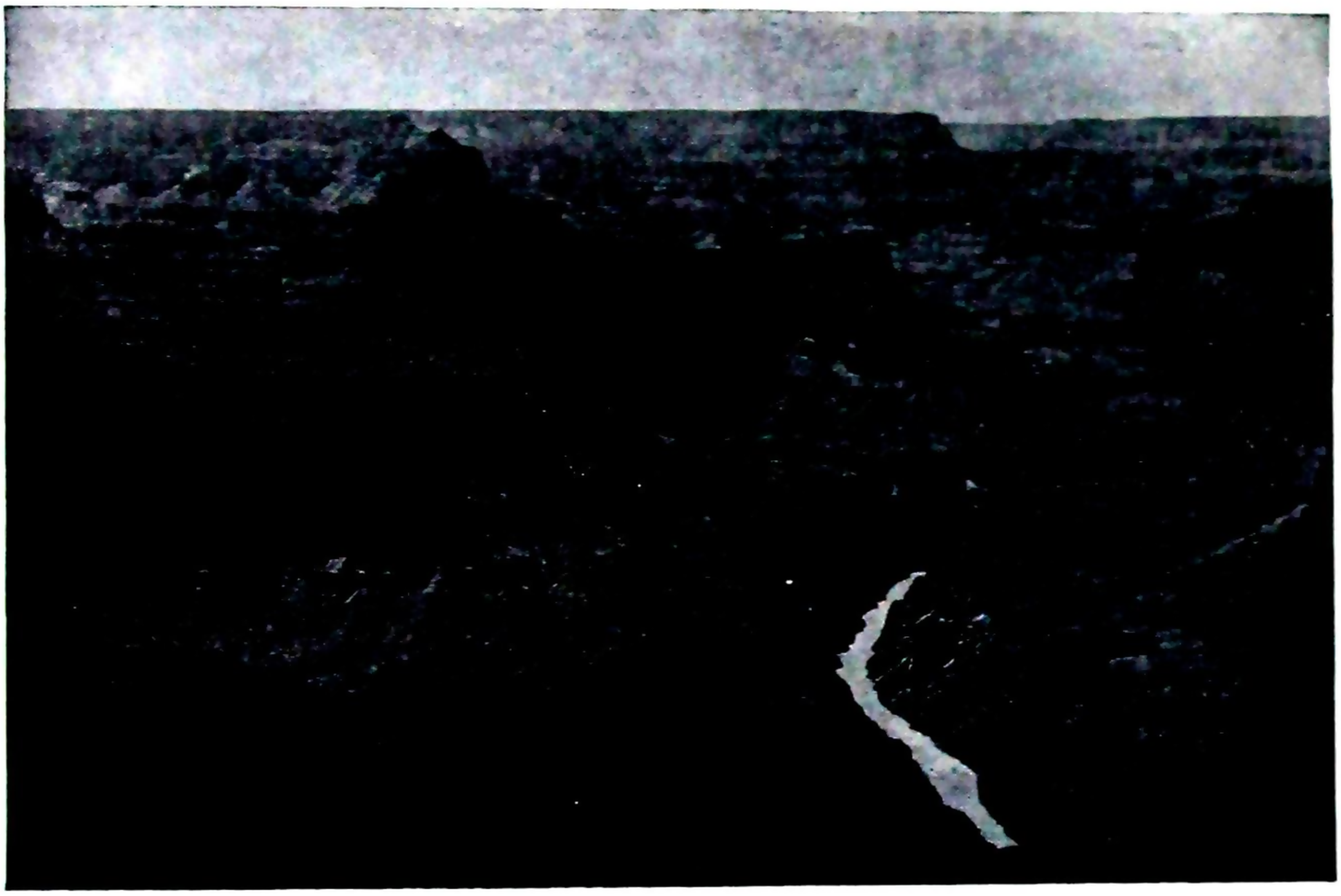


FIG. 49. The Grand Canyon of the Colorado River, a youthful valley. (Photo by George Grant, U. S. Dept. of the Interior.)

Imagine a large, comparatively smooth area that has a fairly uniform slope. As water flows over it from a source at the top, or as precipitation wanders into the lower spots, choosing its own course, it carves out a trench or channel for itself (Fig. 48). The formation of a new river valley has started. The exact location of the channel will depend on the materials which compose the surface and its preexisting irregularities, but its form will be marked by certain definite characteristics. For one thing, at this point in its history the slope is the steepest. The stream will flow rapidly and will perform most of its cutting vertically downward, so that a young valley characteristically is straight and has steep sides and a narrow bottom. The Grand Canyon of the Colorado river, shown in Fig. 49, is an excep-

tionally large example of such a valley. Later on, as the level of the bottom of the valley approaches that of the stream, lake, or ocean into which the river flows, the water moves more slowly, and so does more of its cutting horizontally, widening the valley. The decrease in speed also permits it to be more easily diverted from its straight



FIG. 50. The flood-plain of the Mississippi River showing meanders. (Photo by U. S. Army Air Corps. Pub. by authority Miss. River Commission.)

course. Cross currents may be set up by winds or because of material which falls into the stream from the sides. As a result of these currents, stream banks are eroded irregularly along opposite sides and the river takes on a winding course as the valley approaches maturity. That these curves, descriptively called *meanders*, may become very pronounced is shown in the accompanying photograph (Fig. 50). Places are known in the United States where a boat may travel as much as 30 miles around such a meander and be only one mile from its starting point. This narrow neck may be cut off in flood time and a new channel be established leaving an isolated hill. The net effect of these operations is continually to widen the valley until in old age it becomes a broad flood-plain.

So-called erosional mountains are formed by running water in the manner just discussed. A plateau once existed at approximately the present level of the tops of such a group of mountains or hills. In the centuries since that time rivers and streams have carved out and carried away large quantities of the rock, forming great valleys above which the remaining portions may tower as mountains. The Catskills, Alleghenies, and southern Ohio hills are all examples of such formations.

That part of the rainfall which neither runs off rapidly nor evaporates quickly soaks into the earth to become what is known as ground water. Although most of it soon returns to the air either directly from the soil or through plant transpiration, an appreciable amount makes its way beneath the surface. This water saturates a layer of the earth, the top of which we call the water table. In amount it is believed to be sufficient to form a shell having an average thickness of about 100 feet. Most of the erosional work of ground water is chemical in nature. As soon as it enters the earth it begins to dissolve various materials. In addition to substances ordinarily soluble in water, these include some which become soluble under certain conditions of increased temperature and pressure which may be encountered. Material is removed from one place, transported and later deposited somewhere else. In this fashion huge caves may be formed as, for example, the Carlsbad Caverns of New Mexico and Mammoth Cave in Kentucky. Subsequent deposit of the material results from such causes as lowering of temperature, relief of pressure, or precipitation because of mingling with other solutions. Deposits made in this fashion in rock crevices may result in valuable mineral veins, petrifaction of plant and animal remains, or other formations. Under conditions such that the underground flow of water is sufficiently rapid some mechanical erosional work is also performed.

The wearing away of land by oceans is appreciable, although at present less extensive than that performed by running water or ground water. Such action is largely mechanical in nature, most of it being performed by waves breaking on the shore. When this occurs the water is violently agitated, causing pebbles and sand to grind together and act as tools to wear away the shore or cliff. Shore currents develop which aid in the weathering action. Tides also play a role, al-

though a small one, since they rise and fall comparatively slowly in most localities. Striking examples of the work of waves are found in Great Britain, where some of the chalk cliffs of Dover have been receding at rates as high as 15 feet per year, and where parts of the Yorkshire coast have been worn away a distance of 2 miles since the Roman conquest. The island of Heligoland in the North Sea is another place for which we have the record of the erosional action of waves. This island was reduced in circumference from some 120 miles in 800 A.D. to about 8 miles in 1650 and finally to 3 miles in 1900, at which time Germany fortified it and made it a naval base.

Glaciers

Glaciers constitute another important though temporary agent of erosion. The necessary and sufficient condition for the formation of a glacier is simply that, on the average, more snow shall fall during part of the year than can melt during the remainder so that an accumulation of snow and ice is built up over a period of years. If this takes place over a large, comparatively level area, a so-called *continental* glacier results such as the one now found in Greenland. If the snowy dividends collect in a region of mountains and steep slopes, rivers of ice called *valley* glaciers form and move to lower levels under the influence of gravity. These perform important erosional work. Because of the rocks and other material frozen into the bottom of such a glacier, it acts exactly like a huge rasp or file as it moves across the surface of the earth. Valleys are made U-shaped, and large amounts of earth material are carried along as glaciers of this type proceed on their way. They continue their travels until they either reach the sea or encounter temperatures sufficiently high to cause them to melt. In the latter case the accumulated load of rock and other debris is piled up at the point where the rate of the glacier's advance is just balanced by the rate of melting. Such a deposit is called a *terminal moraine*. Long Island has a geological history of this sort. When a continental glacier disappears because of change in climate, the debris is spread out more or less evenly over a large area. Such a layer of glacial material covers most of the northern middle-western part of the United States and all of New York and New England.

Other Erosional Agents

Factors of less importance include the wind, plants, and chemical effects of the atmosphere. In general we do not think of the wind as being powerful enough to perform any great amount of mechanical work, although when it reaches hurricane proportions pressures from 50 to 100 pounds per square foot may be exerted. It works chiefly

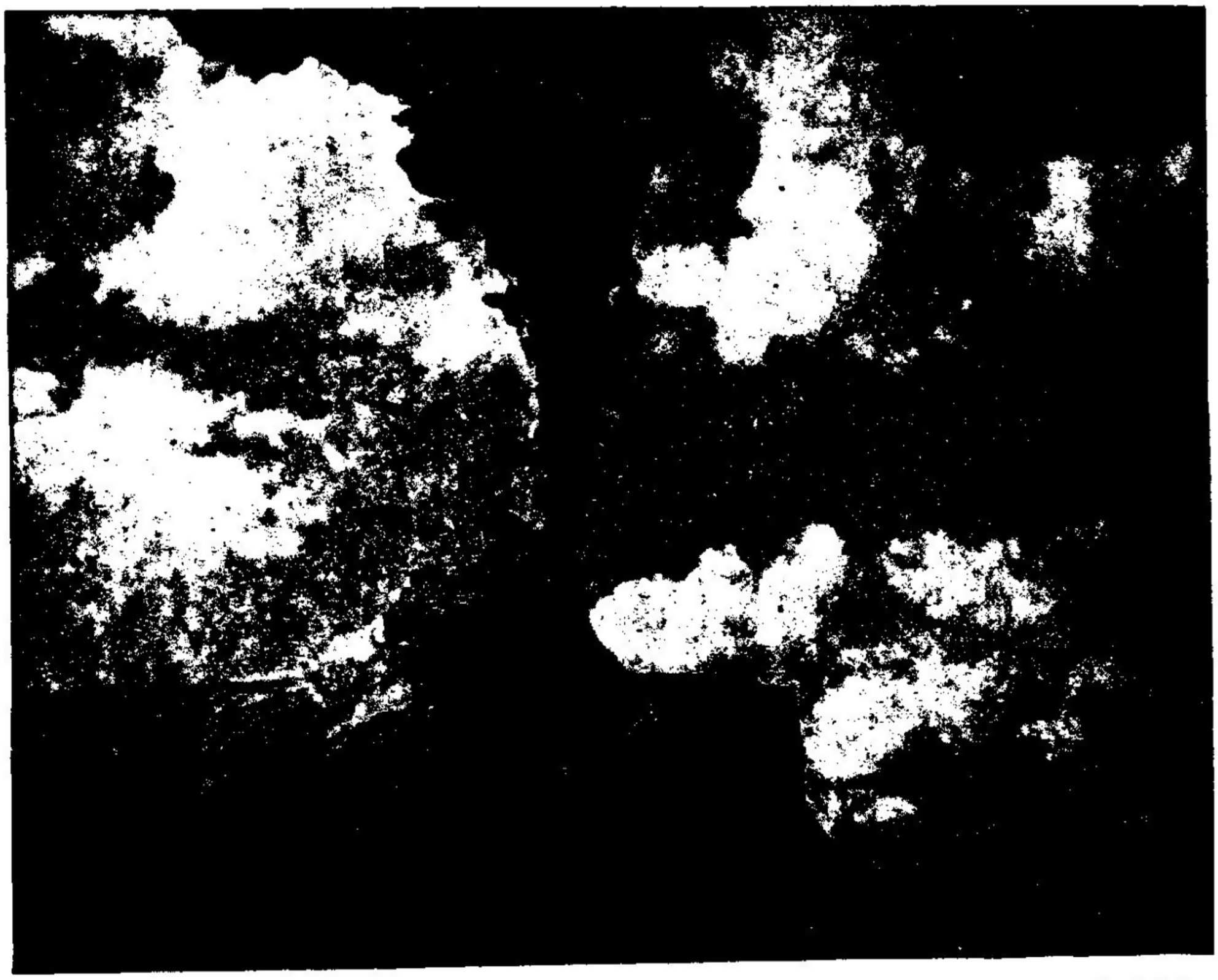


FIG. 51. An example of a structure formed by wind-erosion, Petrified National Forest, Arizona. (Lowell Stagner.)

by picking up loose materials and using these as abrasives to grind up and wear away surfaces against and over which it blows. In other words, it performs in exactly the same manner as the commercial sand blast which man uses, though much more slowly. An example of wind-erosion is shown in Fig. 51. Another contribution which the wind makes to varying the appearance of the face of the earth is the action it has on loose sand. This material, under the action of the wind, drifts like snow and piles up at obstructions to form what are known as sand dunes. In regions where there are prevailing winds, these dunes migrate because the sand is blown up the windward side

and across the top, after which it rolls down the leeward side. Rates of travel are usually less than 25 feet per year, but anything smaller than the dune which is in the way is relentlessly engulfed by it.

Plants add their bit to the general process of making little stones out of big ones. Their weathering work is principally the result of roots creeping into rock crevices, expanding as they grow and thus cracking off pieces of the rock structure. Very large forces may be set up in such cases. Garden pea roots, for example, are capable of exerting pressures as high as 300 pounds per square inch. Most people are familiar with this sort of phenomenon in connection with the havoc often played by roots with tile drains.

The chemical effects of the atmosphere are principally carbonation and oxidation. The carbon dioxide in the air unites with water to form carbonic acid, a weak acid which acts on rock. Oxygen in the air tends to unite with many substances so that oxidation of rock materials aids in the decomposition processes.

All of the above agents work toward the general end of levelling off the surface of the earth. It has been estimated that if they were to proceed unhindered at their present rates, continents would be brought down to sea level in from 50 to 100 million years. Next we will turn our attention to the other group of factors, those which tend to counteract the work of the ones just discussed.

Relief-Building Factors

Isostatic Movements

First it should be noted that these movements with the formidable name are agents which are not seen directly but which are assumed to exist in order to account for certain observed facts. These facts are of two kinds, the first of which concerns the work of running water discussed above. It will be recalled that the Mississippi River was mentioned as depositing in the Gulf of Mexico more than a million tons of mud and silt every day. In one year this would amount to enough to cover an area of one square mile to a depth of more than 260 feet. Adding to this the similar work performed by all the other thousands of rivers and streams on the earth, it is evident that each year enormous masses of material are transferred from high land areas to low land areas and ocean basins. It is also true that this has

been going on for centuries and that as far as the surface of the earth is concerned, there are no agents which take these deposits back again. In other words, in Fig. 52 in which two conical plugs from the surface to the center of the earth are shown, it would seem that plugs like (a) taken from an ocean basin would be getting heavier all the time, while those like (b), from land areas, would grow increasingly lighter. Thus, from one kind of observed fact we reach the conclusion that plugs of the sort described should vary greatly and increasingly in weight.

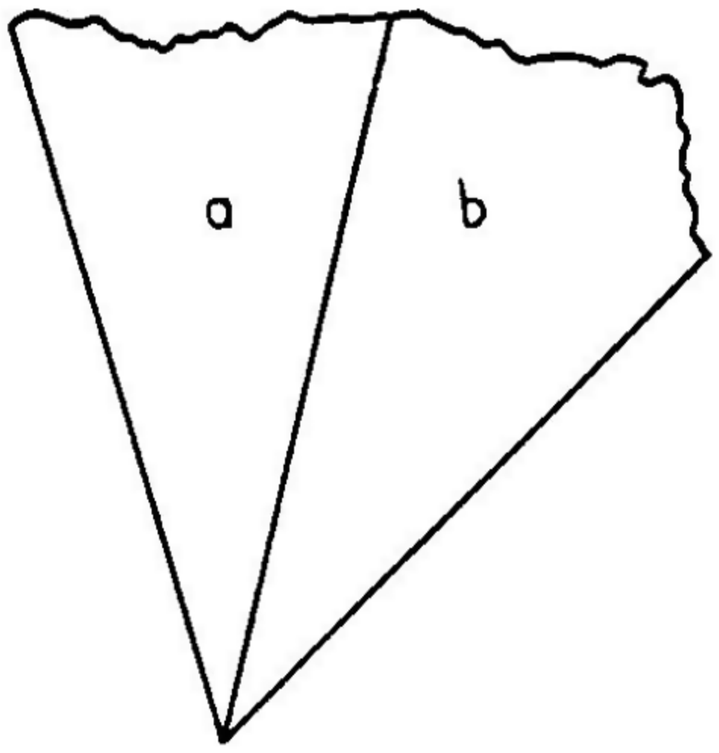


FIG. 52. Diagram showing two conical plugs taken out of the earth, one under the sea and the other under a mountainous region.

Now let us consider the second of the two kinds of experimental evidence. It happens that it is possible to perform pendulum experiments as a result of which the relative weights of these sections of the earth can be determined. Many such investigations have been carried out and the results have consistently indicated that the earth plugs all weigh approximately the same, regardless of the part of the world from which they are taken. This condition is called *isostasy* which means literally " standing the same," and the earth is said to give evidence of always tending toward a state of isostasy. The first conclusion that we may draw from these experiments is that the material beneath the oceans has a greater density, i.e., weighs more per cubic foot, than that beneath the dry land where the plugs are longer. Then we may ask how these two experimental facts can be reconciled. On the one hand we see a constant one-way flow of material from one kind of region to another, while on the other hand the two regions seem to maintain equal weights. Obviously the two can be harmonized satisfactorily by assuming a flow of earth materials at considerable depths below the surface, from regions under ocean basins to regions beneath dry land. These transfers are called *isostatic movements* and are the movements as a result of which the earth is believed to maintain its condition of approximate isostasy. In this way, says the theory of isostasy, as ocean basins become increasingly weighted down and continents relatively lighter because of the erosional work

of running water, there comes a time when the ocean bottoms sink This causes rock materials to flow laterally to regions beneath conti· nental areas, raising parts of them up and equalizing pressures again

There are various surface phenomena which find a ready explana· tion on the basis of this theory. The simultaneous sinking of ocean



FIG. 53. An example of rock folding, Alberta. (W. O. Thompson.)

basins and rise of adjacent continental areas, for example, would be bound to cause disruptions of various sorts in rock layers beneath the surface. From this there might quite reasonably be expected to re· sult the phenomena which we call folding and faulting. In the former, as the name implies, layers of rock are folded or wrinkled, as shown in Fig. 53. Faulting, such as is pictured in Fig. 54 occurs when rock on one side of a crack or fissure moves with respect to that on the other side. Such displacements may be in any direction and vary from less than an inch to several miles. Vertical faulting may give rise to great cliffs. Those along the east side of the Sierra Nevada Moun· tains, extending to heights of several thousand feet were formed in

this fashion. The large mountain chains of the world such as the Alps, the Rockies, the Apalachians, the Himalayas, and the Andes, were formed by folding and faulting.

Further substantiation of isostatic movements comes from various continental areas that are rising or sinking. Elevation is known to be occurring in the Great Lakes region along the Canadian border, exactly the sort of thing that would be expected to occur as a result of the decrease in pressure which followed the disappearance of the last ice sheet from this region some 20,000 years ago. Other cases include northern Scandinavia which is rising, and parts of the southern coast of Greenland which are sinking. The existence of so-called "drowned" river valleys also provide evidence of these crustal movements. Some of the fjords of Norway are glaciated river valleys which were carved out in the usual manner and then at some later time sank beneath the sea. The same is true of part of the valley of the Hudson River which can be traced out into the Atlantic Ocean for some 100 miles beyond Sandy Hook.



FIG. 54. A fault in metamorphic rock near Central City, Colorado. (Worcester's *Geomorphology*, D. Van Nostrand Co., 1939.)

Earthquakes

It has been said that of all nature's spectacular phenomena, earthquakes are the most terrifying in their effect on man. They seem to violate all his ideas of stability and trustworthiness that are implied in the term *terra firma* and leave him with nothing upon which to depend. Therefore, it was quite natural that for centuries earthquakes were accepted as expressions of the awful wrath of an angry God,

who used this impressive means of warning or punishing those with whom He was displeased. It should be noted, perhaps, that there was usually considerable disagreement among men as to just which people were being thus chastised. A few early attempts were made to explain the tremors as due to natural causes but with little success. Among the Greeks, Democritus taught that they were due to heavy rains which soaked the earth and caused it to swell, while Anaximenes attributed them to prolonged droughts which cause the earth to dry out and crack open.

The modern science of earthquakes, *seismology*, which has had almost its entire growth within the last 100 years, is based on the observation that the phenomenon simply consist of waves due to natural causes and traveling in the earth's rocky shell. The causes are believed to be very much the same as those procedures by which man produces waves in the laboratory. There, waves are set up by friction as when a violin string is bowed, by a sharp blow as when a gong is struck, or by the sudden snapping of a substance which has been under great tension. The first two of these are thought to enter to some extent into the production of minor tremors, but major earthquakes are believed to be due to causes analogous to the last of the three. Under the great strains to which rock layers are subjected as isostatic movements proceed, deformation occurs. As tension increases still further there comes a time when the rock gives way, causing the resulting waves and rapid readjustments on the surface which constitute what we call an earthquake. This idea is substantiated by a study of the regions in which the phenomena occur most frequently. As will be seen from Fig. 55 these are belts which border ocean areas, the portions of the earth where surface effects of isostatic movements would most be expected.

Accompanying a major earthquake there is always displacement either along fissures or faults that are already present in the bedrock, or along ones that the quake itself opens. Such displacement may be both vertical and horizontal with the latter usually being the larger. In the San Francisco earthquake of 1906 displacement occurred along an existing rock fracture which has been traced for several hundred miles. During the quake, land on the west side of this fault moved north with respect to that on the east side as much as 20 feet in some

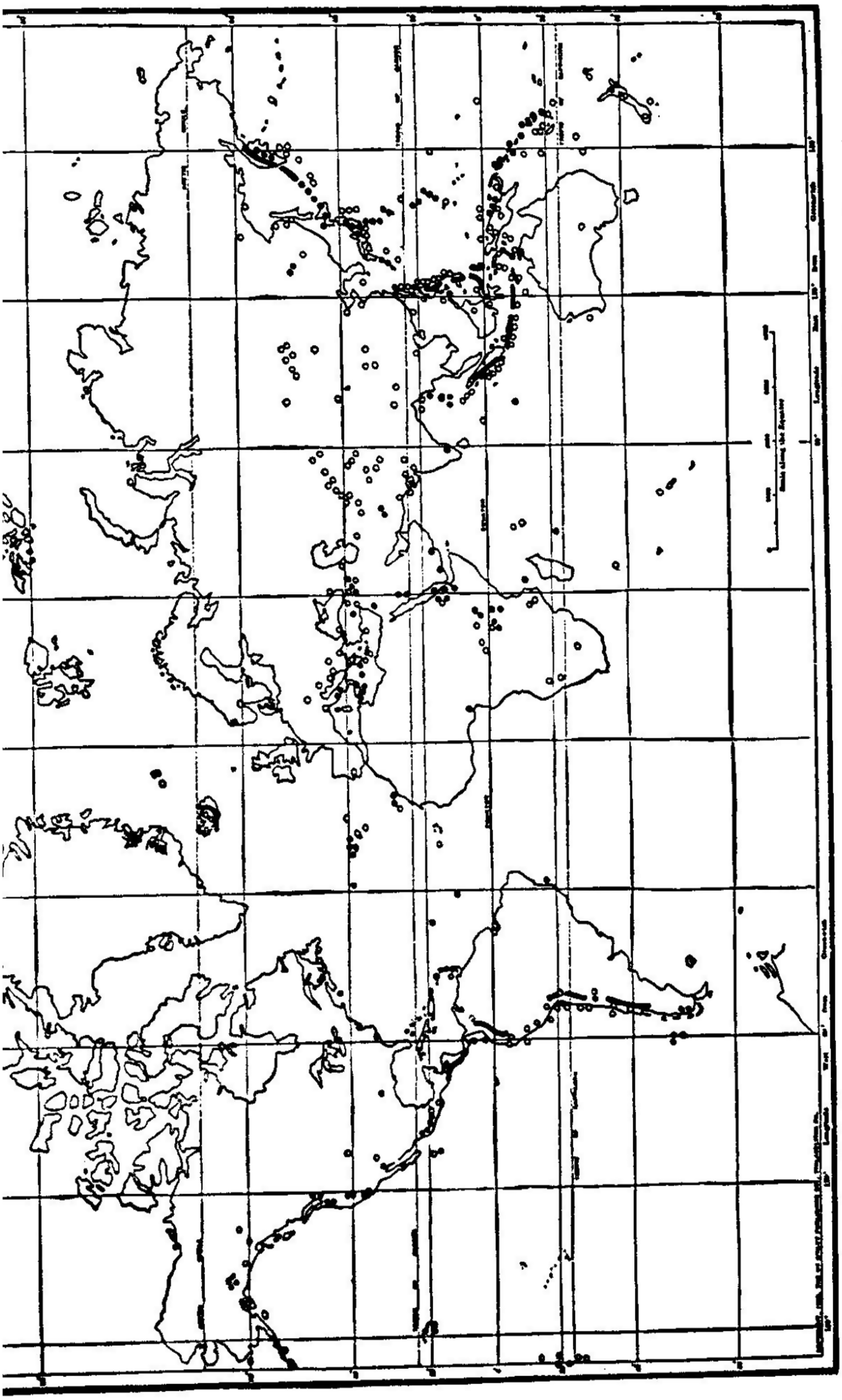


FIG. 55. Earthquake and volcanic regions of the earth. The open circles show earthquake centers. The black circles and solid black lines denote regions of present or recent volcanism. (After Dr. Max Fritz. By permission, from Freitag & Berndt's Tectonische Schulwandkarte der Erde.)

places, severing water mains, gas lines, roads, and fences for many miles.

Several thousand earthquakes of sufficient intensity to register on modern instruments are recorded every year. Most of them occur in the belts mentioned above. Japan averages about four a day the year round and California some 200 each year, the great majority being so slight as to be noticeable only with instruments. The duration even of major earthquakes is seldom more than two minutes with the average being much less than one minute. The 1906 California quake, for example, was of about average duration. Messina, Italy, was entirely destroyed by one which lasted but 35 seconds.

The instrument which registers the passage of earthquake tremors is called a *seismograph*. A modern seismograph is a highly refined and quite complicated device, but the greatly simplified instrument shown in Fig. 56 will serve to illustrate the principles involved. Here we have a solid base which is attached rigidly to bedrock. Upon it is mounted a revolving drum which carries a sheet of paper. Suspended from the cross arm is a heavy weight to which is attached a stylus adjusted to trace a line on the paper. As long as stylus and revolving drum do not move vertically with respect to each other, the line will be straight. But when an earthquake tremor passes through the bedrock, the vertical component of the wave will cause the base, and therefore the drum, to move up and down. The suspended weight, since it hangs from a spring, will remain approximately stationary and a wavy line will be traced on the paper. In practice a second instrument arranged to react to horizontal waves would also be employed.

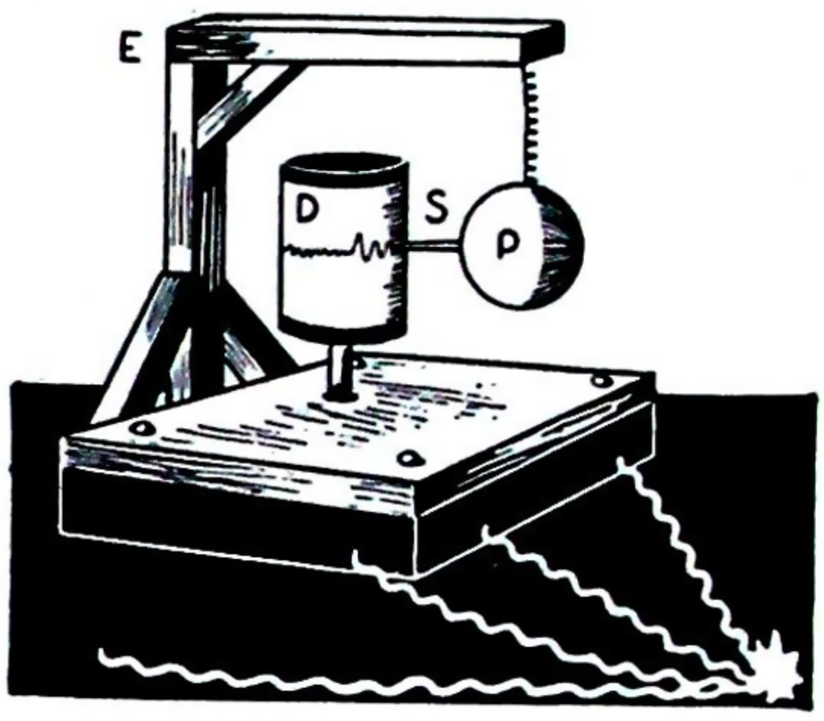


FIG. 56. Diagram of a simple seismograph. (After Croneis & Krumbein's *Down to Earth*, University of Chicago Press.)

Seismograph records show that traveling out from the typical earthquake there are three distinct kinds of waves. The first one to reach the seismograph goes directly through the earth. It is a longitudinal or compressional [1] type wave and may travel as rapidly as 300 to 400 miles per minute. The second one also goes directly through

[1] See Chapter 28, pages 537–538, for definitions of *types of waves*.

the earth but somewhat less rapidly. This one is of the transverse variety. The third wave is the slowest of the group, but since it travels around the outside of the earth, it is the most destructive.

We commonly think of earthquakes almost entirely in terms of the damage they do to us and to the little structures we erect on the surface of the earth. As a matter of fact, through the seismograph records they make, earthquakes are of considerable value. One of the uses to which such data are put is the determination of the center of the earthquake disturbance. It has been shown to be true that the time between the arrival of the first and second waves depends only on how far away the quake occurred. Since with modern seismographs it is possible to determine accurately the direction from which the waves come, the exact location of the disturbance may be found.

From these data also we may learn certain things about the interior of the earth. The velocity with which a wave travels through any substance depends on the density and elasticity of that material. The quantitative relationships involved have been worked out carefully and accurately in the laboratory. Now in reaching different seismographs scattered about the earth's surface, the waves from a given quake will travel through materials at varying depths beneath the surface. For this reason, data on the velocities of the waves make it possible to determine the density and elasticity properties of the substances through which they passed. It is principally upon investigations of this sort that the conclusions about the interior of the earth, mentioned in the preceding chapter, are based.

Vulcanism

Magma or molten rock was mentioned briefly earlier in the chapter in connection with the discussion of igneous rock. Now we are concerned with various effects it has on the contours of the earth's surface. Different types of igneous formations are illustrated in Fig. 57.

When magma is forced in between rock layers without reaching the surface, so-called igneous intrusions result. These are of several kinds. The magma may simply flow between horizontal layers of sedimentary rock and solidify in a large flat sheet of igneous rock called a sill. Such a sill composes the structure known as the Palisades of

the Hudson River. Or it may find its way into cracks and faults which cut through sedimentary layers. In such a case the solidified rock sheets will be more or less vertical and if subsequently exposed by erosion have the appearance of a huge stone wall. These are called dikes (Fig. 58) and may vary in thickness from hundreds of feet down to paper thinness. Or the intrusion may take place with sufficient pressure to bow up the layers of rock above it, forming hills or

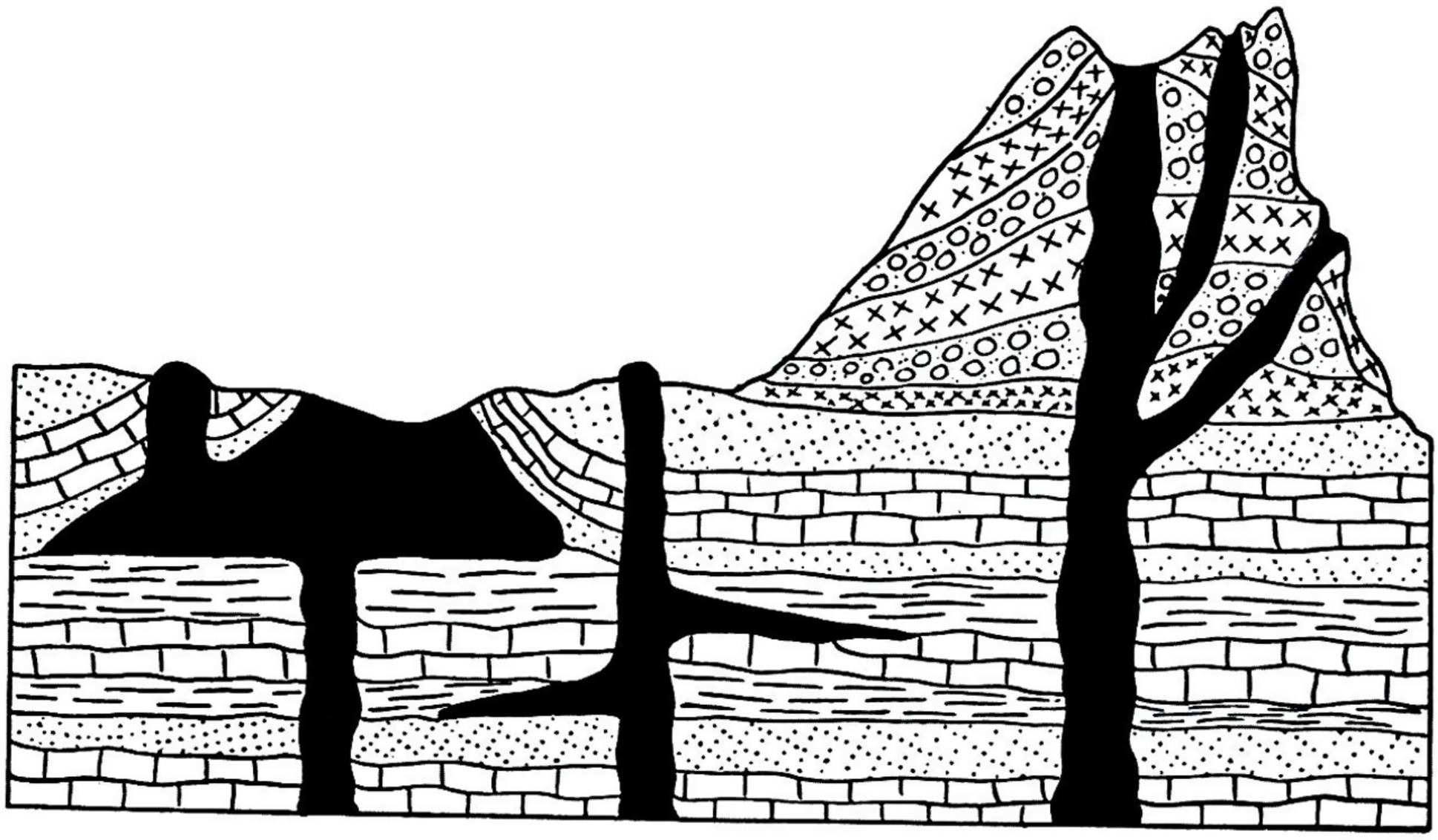


FIG. 57. Diagrammatic representation of several important phases of vulcanism. At right is a volcanic cone built by alternating lava flows and layers of volcanic ash. The center shows an extinct volcano whose cone has been eroded away. Note hardened lava projecting above the surface as a volcanic plug. Two sills are shown. At the left is shown an intrusion which domed the beds above it. At the extreme left is represented a dike. (After a drawing in *Earth History* by L. C. Snider, The Century Co., 1932.)

so-called dome mountains on the surface of the earth. The Henry Mountains of Utah and the Black Hills of South Dakota were formed in this fashion.

Of the various formations which result when magma finds its way to the surface, volcanoes are the best known. Volcanic eruptions vary all the way from slow quiet outflows of lava to what are essentially explosions of great violence. In the former case an inverted saucer-shaped mountain is built up by the hardening of successive layers of lava one above another. The cone of Mauna Loa is typical of such a volcano. Its comparatively symmetrical form rises

14,000 feet above sea level. In fact the entire chain of the Hawaiian Islands was formed in this manner, rising to a maximum height of 36,000 feet above the floor of the Pacific Ocean. At the other extreme are the eruptions of explosive violence. In these cases great clouds of ash and volcanic fragments may be thrown out, accompanied by little lava flow. The classic example of this occurred on the

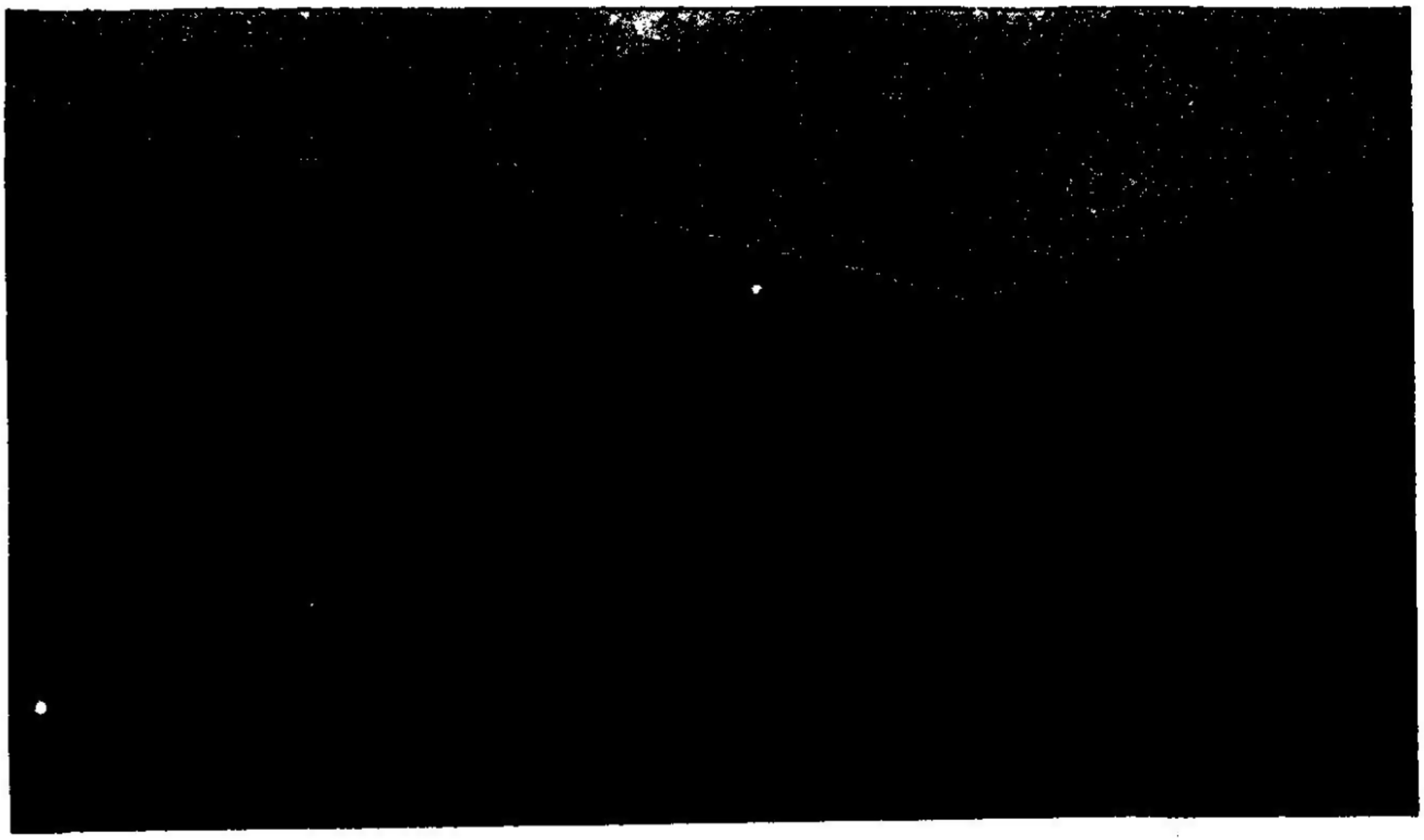


FIG. 58. One of the great Spanish Peaks-dikes, Colorado. Horizontal lines on the face of the dike indicate former position of sedimentary rocks which have been eroded away. (U. S. Geol. Survey.)

island of Krakatoa in the Indian Ocean in 1883. Here with almost no warning the island was literally blown to pieces. The noise of the explosion was heard for 3,000 miles and dust from it was carried by the atmosphere over the entire earth.

Occasionally, instead of building large cones about a very few openings, exceptionally fluid lava will emerge from a large number of vents and spread out more or less evenly over a large territory. Such flows occurred repeatedly during the past few hundred thousand years (the most recent 8 to 10 thousand years ago) in the region we now know as the Columbia plateau in the northwestern part of the United States. More than 100,000 square miles of Washington, Oregon, Idaho, and northern California are covered with this sheet of lava which varies in thickness from 10 feet to several thousand. The only such lava flow of any extent in historic times occurred in Iceland about the time of the American Revolution. Fissures opened

up joining together the craters of three volcanoes, and lava estimated at some 500 cubic miles flowed over the surrounding country within a period of about six months. One entire end of the island was completely altered in topography, being changed from a hilly region to one consisting of a high plateau.

Still another kind of formation may result when the lava which fills the neck of a volcano cools and hardens there as the volcano itself becomes extinct. If the softer surrounding parts of the cone are eroded away at some later time, this huge, more or less cylindrical shaft of igneous rock is left exposed. Such a structure is called a volcanic plug. Devil's Tower in Wyoming which rises some 600 feet above the surrounding territory had such an origin.

The map shown in Fig. 55 indicates the major volcanic belts on the earth. Note that they, like earthquake belts, are concentrated along shore lines to a considerable extent.

So we see that there are a number of agents constantly at work altering the appearance of the face of the earth. In the next two chapters, after having considered briefly the bases of calculation of the age of the earth, we shall examine the sorts of clues that man looks for in studying the earth's history and then see some of the things he has learned about that history.

Questions

Discussion

1. Define the several types of rock and give an example of each.
2. Look up some information about one or two of the earlier theories that were advanced to explain the origin of the heat which melts rock beneath the earth's surface.
3. What are the approximate average percentages of iron and oxygen in igneous rock?
4. Describe a sequence of conditions under which a given mass of rock materials might change from sedimentary to metamorphic to igneous and back to sedimentary again.
5. Define " relief-levelling " and " relief-building " agents.
6. The work of which one of the various agents discussed would differ the most from what it is now if rock did not expand as it increased in temperature? Explain.
7. Describe the life history of a typical river valley.
8. Give several reasons for believing that rivers carve out their own valleys.

9. What are "meanders" and what relation do they have to the formation of river valleys?

10. Look up one or two specific examples of erosion by waves, other than those mentioned in this chapter.

11. Look up some information regarding the glacial history, if any, of your locality.

12. What is the necessary and sufficient condition for the formation of a glacier in any region?

13. Describe the experimental evidence on which the theory of isostasy is based.

14. Mention two observed facts, other than the ones given in answer to the preceding question, which are in agreement with the theory of isostasy.

15. Distinguish between folding and faulting.

16. Describe a typical seismograph record and show what use can be made of it.

17. What experiment, discussed in an earlier chapter, gives results concerning the nature of the earth's interior which are in agreement with those obtained from earthquake data?

18. Identify the following: sill; dike; magma; volcanic plug; dome mountain.

19. Look up additional information about the destruction of the island of Krakatoa in 1883.

Multiple Choice

1. The record of an earthquake, as recorded by a seismograph, consists of three waves, (a) two of which travel directly through the earth, (b) all of the same velocity and amplitude, (c) the most rapid of which causes the most destruction, (d) all of which enter into the scientists' calculation of the location of the quake, (e) only one of which goes through the earth instead of around its surface.

2. Magma which cools and hardens beneath the surface of the earth is called (a) lava, (b) conglomerate, (c) metamorphic rock, (d) mantle rock, (e) igneous rock.

3. If rock were a very good conductor of heat, the amount of erosion due to which one of the following factors would differ the most from what it is now? (a) Running water. (b) Glaciers. (c) Ground water. (d) Temperature variations. (e) Freezing and thawing.

4. Which one of the following seems to offer the most satisfactory explanation of the observed rise of the Great Lakes region? (a) Weathering by running water. (b) Isostatic movements. (c) Volcanic deposition. (d) Variation in carbon dioxide content of the atmosphere. (e) Earthquake tremors.

5. Pendulum experiments performed at different points on the earth's surface give results which are most contradictory to which one of the fol-

lowing statements? (a) Running water deposits heavier materials first and lighter ones later. (b) Oceans originally consisted of fresh water. (c) An : foot of the materials which underlie the oceans weighs less ..... much as an average cubic foot of that under mountainous regions. (d) Earthquakes are caused by drying and cracking of the lithosphere. (e) In times past, the oceans have covered a larger percentage of the earth's surface than is now submerged.

True–False

1. Most earthquakes occur in regions where isostatic movements would be most expected to take place.

2. Geological evidence seems to indicate that throughout all of earth history, the forces of gradation have had the upper hand over the relief-building group.

3. If a valley glacier melts at the same rate as it advances, the debris it deposits will be spread out evenly over a large area.

4. If a valley has steep sides and contains a swiftly flowing stream, it is probably comparatively old.

5. Fossils are usually found in marble or slate deposits.

6. Basalt and granite are both forms of igneous rock.

7. In order to obtain the best results, a seismograph should be mounted on a soft rubber base.

8. A so-called "volcanic plug" results when the sides of a volcano cave in, filling the neck and preventing further eruption.

9. The peaks of the Rocky Mountains mark the level of what was once a huge plateau, most of which was eroded away leaving the mountains exposed.

10. Igneous rock composes over 90 per cent of the earth's crust.

Chapter 9

CLUES TO EARTH'S HISTORY

Sir James Jeans has said that if the Woolworth Tower were to be thought of as standing for geologic time, the length of human existence could be represented by a nickel lying on the tower, and historic time by a sheet of paper on the nickel. Therefore it is obvious that any knowledge we acquire about the changes that have occurred on the surface of the earth since its beginning must be gained by some other means than direct observation by human beings. As a matter of fact, the only record man has of the earth's history is the one it, itself, has written in the rocky shell about it. In reading this record the geologist is in much the same position as the detective working on a murder case. He must search for clues, set up tentative hypotheses from the clues, check and compare the hypotheses with each other, and in this fashion finally reach conclusions. These clues upon which the earth-detective depends will be our principal concern in this chapter.

First, however, there are two introductory matters that must be mentioned. For one thing, the reader should have in mind a clear picture of the planetesimal hypothesis of the formation of the solar system, since in general it provides the language in which the earth's beginning is discussed today. For this he may look back at the discussion in Chapter 3. In the second place it seems desirable to give a little consideration to the evidence for the very great age which we ascribe to this our favorite planet.

The Age of the Earth

We have already referred to the fact that as recently as a century ago it was a widely accepted belief that the earth had been in existence only a few thousand years. This belief was chiefly due to a declara-

tion made by Archbishop Ussher of Ireland about the middle of the seventeenth century. On the basis of his interpretation of certain Biblical passages, the good Archbishop made some calculations and confidently announced to the world that the planet had been created on October 26, 4004 B.C. at 9 o'clock in the morning. This was immediately adopted by the church as being correct, and so it is not surprising that people in general were very slow to abandon it when later scientific evidence began to contradict it. Incidentally, it should be pointed out that the date mentioned was based on the Archbishop's *interpretation* of certain scriptural passages and therefore was his responsibility, not the Bible's. In other words the fact that we now know the earth cannot possibly be so young should not be held against the Bible anymore than the fact that Laplace misinterpreted the nature of the rings of Saturn is taken as a reflection on the good name of that planet.

Something like forty more or less scientific approaches have been made to the problem of determining the age of the earth. We will consider briefly three of these, two of which are now looked upon as quite unsatisfactory and one which as far as present evidence goes seems to be reliably founded. One of the earliest attempts to solve the problem assumed that the oceans were originally fresh water and then proceeded to calculate the time that would be necessary for them to acquire their present salt content. A simple division of the amount of salt now present by the average annual rate at which it has been carried to the sea through the ages would, of course, give a figure for the time the oceans have been in existence. This could then be taken as a minimum age for the earth. The difficulty lies in the fact that there is no way of determining the average rate at which salt has been carried into the seas, nor of correcting for loss of salt from the sea during all the years. The present rate of deposit is certainly much greater than it has been through much of the earth's existence, since during most of the time the slopes down which water flowed to the sea were much less steep than at present. This means that any figure obtained by this method is far too low. Calculations made in this manner gave an age of something like 100 million years.

Another method is based on the rate at which sedimentary rock deposits were built up. Again, the difficulty is that there is no reliable way of determining what rates of deposit have prevailed in the past.

Using present rates, the method has given results varying from 100 to 300 million years which, as in the previous case, may be presumed to be too low. Present rates of deposit are unquestionably higher than an all-time average would be. In addition to this, the method takes no account of the periods of erosion between deposition phases.

The method of determining the earth's age which science today accepts as the most reliable is based on the peculiar behavior of a group of so-called " radioactive " substances. Since these materials will be discussed more fully in a later chapter,[1] a brief word of description here will suffice. Atoms of such elements have the startling habit of occasionally shooting out certain of the fundamental particles of which they are made and in this manner becoming atoms of entirely different elements. This process is known as radioactive disintegration. The effect is somewhat as though a large brick factory building should suddenly shoot out a lot of bricks and become a theater, which after a time would emit more bricks and settle down into a dwelling house, which later on would repeat the procedure and continue its existence as a hamburg stand. In a similar fashion radioactive substances are grouped into series or chains. The heaviest of all the elements, uranium, heads one of these. When an atom of uranium disintegrates it becomes an atom of an element known as uranium X. Uranium X, in turn, decomposes into the third member of the series. This continues on through radium, which is sixth in line, and finally ends with the fifteenth, which is lead. Here the process either stops or becomes too slow for us to detect as yet. The value of the phenomenon as a geological clock lies in the fact that the rates at which the disintegration processes occur have been carefully observed and found to be quite constant for each element. It is known, for example, that some 210 tons of uranium will give rise to one-thousandth of an ounce of lead each year. Consequently, when a rock deposit is found which contains both uranium and lead it is possible to calculate the time necessary for the formation of the latter from the former. At this point the careful reader may ask two questions, first, " What right has one to assume that all the lead present came from uranium? " and second, " What right has one to assume that rates of radioactive disintegration have always been the same as they are today? " With

[1] See Chapter 26, page 484.

regard to the first of these, nature has been very kind, because the lead that comes from uranium has an atomic weight slightly different from that of ordinary lead and so is easily identified. The justification for assuming present rates of radioactive decay to have existed in the past lies in the fact that man himself has been quite unable to change these rates in any way. He has subjected the elements to high and low temperatures, high and low pressures, strong electric and magnetic fields, in fact to every sort of extreme environment he could devise, and all without apparent effect on the speed with which their disintegration programs proceed.

The greatest age that has been obtained to date for a specific rock structure, calculated by the above method, is 1,850,000,000 years, which would give an absolute minimum age for the earth. Various other considerations which have to do with the probable extent of the earth's early infancy before solid rock deposits and radioactive disintegration entered the picture lead geologists to talk in terms of a total age of from 2,000,000,000 to 4,000,000,000 years as a reasonable minimum, with the possibility that it may be much older.

Rock Correlations as Clues to Earth's History

A considerable amount of information about the sequence of geological events in any given region can be obtained from rock layers themselves without regard to their fossil content. Our purposes in this connection can perhaps be served best by considering a number of specific examples. Since the scope of this book makes it neither possible nor desirable to attempt to explore exhaustively the possibilities in rock formations, it must be emphasized that the following constitute simply a few typical illustration and the possibilities they offer as sources of geological information.

Examples of Conclusions Drawn from Rock Formations

We will begin with the fundamental principle that if a series of layers of sedimentary rock are exposed, the uppermost is assumed to have been deposited most recently, the second next most recently and so forth. This seems almost too obvious to need stating and yet, even here, one must not be too hasty in any given case since it sometimes happens that folding completely inverts a series of rock layers.

The kind of rock in a given deposit may be of value as a clue to earth conditions at the time the material was deposited. The presence of an extensive bed of limestone indicates that the region was under water during deposition. One may also infer that the surrounding territory was comparatively low and flat since otherwise streams would have carried in mud and silt or coarser materials. Large limestone deposits also are generally taken to imply the existence of certain forms of marine life, since today beds of this sort are laid down through the agency of such organisms.

Now, if this layer of limestone is covered with a layer of shale or sandstone, it is evident that after the limestone was deposited, lands adjacent to the sea were raised with respect to the ocean bed, causing streams to flow more rapidly and transport earth materials. The reverse of this series of events would be inferred to have taken place if the shale or sandstone bed is just beneath the limestone.

If the two layers mentioned above, namely limestone and shale, were laid down successively without the area having been raised out of the water at any time, one would be deposited evenly upon the other. In other words the plane or surface of division between the two would be smooth and unbroken, a formation which is known as a *conformity*. If, on the other hand, there was a period of time between the deposition of the adjacent layers during which the lower one was out of the water and exposed to erosion, the dividing surface would be uneven and jagged, a phenomenon called an *unconformity* (see Fig. 59). So the existence of a conformity or an unconformity between two rock strata provides information concerning the sequence of events when the layers were deposited. The degree of coarseness of the material in a sedimentary bed is a clue to the sort of slopes that existed in the territory surrounding the region. The steeper the slopes, the more rapid are the rates at which streams flow, and the greater are their capabilities for transporting rock materials.

Glaciers leave behind them various clues as to their general nature and behavior. Grooves are cut in the rock over which one of these rivers of ice passes, providing information as to its direction of travel. The maximum vertical thickness of a valley glacier may be determined by observing the upper limit of the glacial debris along the sides of the valley. The present distribution of rocks carried by the glacier

will indicate something about its type and its fate since, as was suggested in the preceding chapter, a glacier that melts as rapidly as it advances piles up the debris in a great ridge, while one that disappears owing to a change in the climate, deposits its load of rock more or less evenly over a large area. The former size of a continental glacier will of course be indicated by the extent of the region covered with the debris.

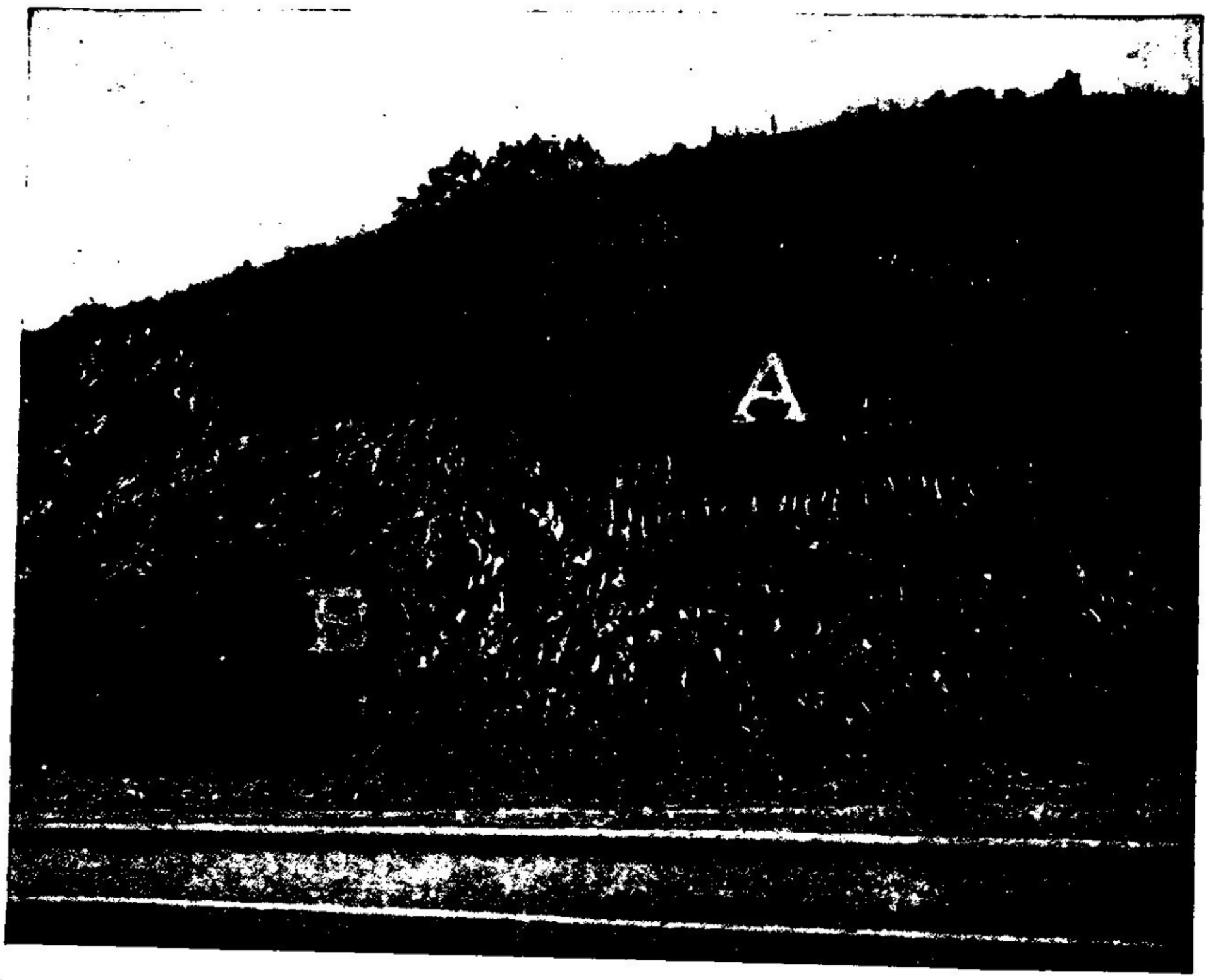


FIG. 59. Nearly horizontal beds of alluvium (A) of Late Cenozoic resting by unconformity upon tilted beds of marine sandstone and shale (B) of Middle Cenozoic. Campus of University of California at Los Angeles. (Miller's *Historical Geology*, D. Van Nostrand Co., 1937.)

Former crustal movements may often be inferred from present positions of rock structures. Consider, for example, the formation shown in (*a*) of Fig. 60, in which are found approximately vertical rock strata overlaid with horizontal layers. The history of such a formation might be as follows. Since all sedimentary rock layers are deposited in a horizontal position, or nearly so, it may be assumed that those which are now vertical, originally were laid down as approximately level beds, as shown in (*b*). As a result of folding, these layers were bowed up into the form shown in (*c*). Then, after a long pe-

riod of being out of the water and therefore subject to the agents of erosion, the structure was worn down to the form seen in (*d*). When at some later time new crustal movements caused the region again to be engulfed by the sea, other sedimentary strata were deposited on top of the weathered vertical ones. At a still later date it emerged from the water a second time and finally was found by man in the form originally pictured in (*a*). Such a series of events could produce this sort of formation. Whether or not any given structure had exactly this history would be determined by careful investigation of all available clues in the vicinity.

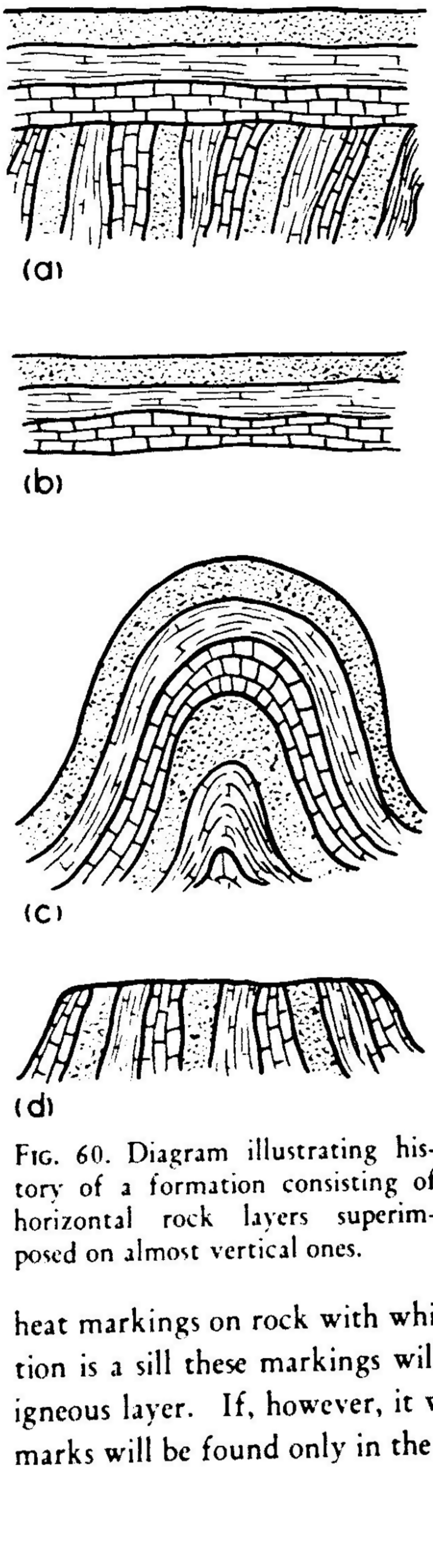


FIG. 60. Diagram illustrating history of a formation consisting of horizontal rock layers superimposed on almost vertical ones.

One more example will conclude our discussion of rock correlation case histories. Suppose a horizontal layer of igneous rock is discovered sandwiched between two layers of sedimentary rock. The problem is to determine whether it was intruded between the two as in the formation of a sill, as discussed previously, or whether it simply flowed out over the surface of the earth and was later covered with a layer of sediment. Careful examination of the rock in contact with the igneous material will settle the question. The high temperature of lava or magma will cause it to leave certain heat markings on rock with which it comes in contact. If the formation is a sill these markings will be found both above and below the igneous layer. If, however, it was formed in the second manner, the marks will be found only in the sedimentary layer below.

Limitations of Rock Correlation Clues

Before considering the second type of clue to earth history, the limitations of the rock correlation method must be mentioned. By procedures of the sort just discussed it is possible to work out a fairly complete picture of the sequence of geological events for any given region. However, any complete picture of the history of the earth as a whole requires that one be able to say, for example, not only that the layers of rock in northern Ohio were deposited in a certain order, but also that any given layer in Ohio was deposited at a certain time with respect to a layer in Canada, or Europe, or some other part of the world. For such purposes rock correlation methods alone are quite inadequate. Rocks, as we have seen, are of only a few types. The rocks of any one type resemble each other very closely regardless of the time they were formed. Limestone which was deposited millions of years ago may not differ in the least from limestone that is being deposited today. The same is true of igneous, sedimentary, and metamorphic strata. For information that will lead to worldwide correlations of geological events we must turn to the other type of clue.

Fossil Remains as Clues to Earth's History

The fossil remains of plant and animal organisms offer the only satisfactory and reliable basis for relating geological events of large areas of the world to any sort of an absolute time scale. It should be noted that the ideal method of dating rock layers would be always to have available igneous intrusions containing radioactive elements. But it happens that 75 percent or more of sedimentary rock strata lack any sort of igneous intrusions.

Kinds of Fossils

In its geological use, the term "fossil" has a somewhat wider application than is usually associated with it. In earth science fossils may be thought of as including "any traces of plants or animals which, having been buried in rock-forming mud or sand, are preserved in such a condition that they give information about the organism or its environment." [2] Within this rather broad definition, fossils may be grouped into several types.

[2] Snider, L. C., *Earth History*, p. 47. New York: Century Co., 1932.

Rarest among fossils are organisms which have been completely preserved. Such remains are found, for the most part, in ice, amber, or oil, since these substances serve to protect the soft parts of the organism. A few hairy mammoths that lived at about the time of the last glacial age have been discovered intact, refrigerated in the ice of Alaska and Siberia. Several were so well preserved that the meat was eaten by dogs, and at least one case has been reported in which steaks from one of these extinct cousins of the modern elephant were served at a scientific dinner. A few members of the wooly rhinoceros family have been discovered preserved in the oil-saturated soil of parts of Poland. Since both mammoth and rhinoceros date back only some thousands of years, they are not of great value from the standpoint of early geological history. The oldest of the completely preserved specimens are insects, many of which have been found encased in amber. Some of these date back as much as 30 or 40 million years.

The unaltered remains of the hard parts of organisms are found much more commonly. These include bones, teeth, tusks, claws, horns, and fins. Important examples of this type of fossil are the dinosaur remains which have been recovered from certain rock deposits in our Great Plains region, and the skeletons of the sabre-toothed tiger, some 800 of which have been found in the asphalt beds of California. These remains are of animals which had their day in the sun a few millions of years ago.

Specimens of any great geological age have usually been altered in some way or other. Plants, for instance, are often preserved by carbonization. This usually occurs when the organism, which may be anything from a leaf to an entire tree, falls into the water and is buried under sediment. During the processes of incomplete decay, under these conditions, hydrogen and oxygen escape, leaving the residue with an ever increasing percentage of carbon. These carbonized remains may retain in fairly exact detail the markings and structures of the original organism. Almost everyone has seen such fossil remains in coal. Then there are a large number of mineralizing or petrifying agents whose action results in the replacement of plant and animal remains by other materials. Here the structure is often reproduced and preserved for hundreds of millions of years in microscopic detail. Petrified wood is an example of such a formation.

Finally there are those things which are evidences of the former existence of living organisms, although never actually a part of them. The shell of a snail may have been buried in a type of sediment which packs tightly about it and hardens. Then if the shell is dissolved out, a mold is left. From such molds casts can be made which reproduce the surface features of the buried object. Sometimes natural processes refill such a mold making a natural cast. Although not important as clues to earth history, a number of spectacular cases of such molds were unearthed in the excavation of Pompeii. Many human beings as well as other animals were caught and completely buried in the lava flow. Careful excavation in modern times has made it possible to prepare casts from the molds thus made in igneous rock.

In this group of fossils also there must be included such formations as dinosaur footprints, stomach stones of certain ancient reptiles, worm burrows, and similar things which provide information about organisms that once lived on the earth.

Abundance of Fossils

In comparison with the great multitudes of living organisms that have existed upon the earth, the number of fossils that man can hope to find is extremely small. In the first place the vast majority of plants and animals, when they die, leave no trace of ever having existed. Most plants decay or are eaten by animals. Most animals are eaten by other animals or decompose soon after death. So that, in order to stand a chance of becoming fossilized, an organism must in general be buried very shortly after it dies. For this reason marine life is more likely to be preserved than land life. It is evident that only a few perish under circumstances such that the remains even can start along the road to " fossildom." Then of these few, many will be destroyed as time goes on. Some will be dissolved out and not replaced. Others will be ruined by the forces of metamorphosis. If sedimentary rock melts and forms magma, any fossils it contains will be destroyed. Finally of the organisms that do become fossilized and remain so, man may expect to find only a small fraction, since he has access to so little of the entire amount of sedimentary rock in the earth's outer shell. Nevertheless, his total fossil discoveries are large and remarkably complete and connected.

It should perhaps be mentioned that only within relatively recent times have fossils been generally accepted as being the remains of organisms that once lived on the earth. Down through the ages many theories were advanced to account for the existence in rock of these interesting specimens. Some people thought they were spontaneously generated. Others believed them to be in the nature of "seconds" which God had discarded during His activities incident to creation. Another theory pictured them as the remains of organisms which perished in Noah's flood. When they were first being used as evidences of evolution, the suggestion was advanced that they had been planted in the earth by the Almighty in order to test man's faith. It was really only about the beginning of the nineteenth century that their true nature began to be fairly widely recognized.

The Use of Fossils in Dating Rock Layers

Modern use of fossil records as clues to geological history is based on the fundamental principle that if fossils of a certain organism are found in a layer of sedimentary rock, then that layer was deposited at the time that particular organism was in existence on the earth.

In using fossils as a basis for fixing the sequence of deposition of rock strata in various parts of the earth, one begins with a region in which the layers of rock lie flat, one above the other, and have not been distorted by folding. In such a locality the fossils in as many layers as possible are carefully studied and classified. Then with regard to each stratum one determines the kind of rock it is, the kinds of fossils that are peculiar to it, and the ones which are most abundant. These records then serve more or less as standards with which rock strata in other regions can be compared and consequently dated. By working in this manner from place to place a geologist can build up general sequences which simplify the dating of new regions.

In order to give one or two examples it will be necessary to anticipate the next chapter, to the extent of mentioning that well-connected geological history is divided by man into three eras, which for the time being we will designate simply as I, II, and III. Now it happens that no mammal bones have ever been found in rocks of era I and only a few of small mammals in rocks of era II. Therefore

whenever bones of large animals of this type are found in rock strata, those layers are recognized as belonging to era III. Similarly large numbers of fossils of certain crablike animals called trilobites have been found in rocks of era I but none have ever turned up in era II or era III strata. Consequently, the presence of a fossilized trilobite in a sedimentary bed dates it as belonging to the first of the three divisions. Certain other marine organisms, on the other hand, are found only in rock layers of eras II and III and so serve to date those beds.

Lest this whole procedure sound somewhat arbitrary it should be emphasized that in all the investigations that have been made, the record is absolutely consistent. It will be evident even from the above few examples, that the discovery of a large mammal bone in a lower rock layer or even in the same layer with trilobite fossils would completely discredit the whole procedure. But no such case has ever been found in all the thousands of locations that have been studied. The same is true of all the other forms which enter into this use of fossil remains. Since this complete coherence does exist, one feels justified in assuming the correctness and reliability of the method.

One other point that may be mentioned in this connection is the fact that the use of fossils as indicators of rock age assumes that plants and animals have changed with time. Suppose that sabre-toothed tigers had always existed, were still roaming the earth, and had not changed in form. Then the discovery of the remains of one in a rock bed would mean nothing in terms of dating that stratum. The very fact that change has occurred, that new forms have appeared and others become extinct, is the thing that makes the procedure of rock dating possible.

Fossils also are good clues to former climatic and topographical conditions, since certain types of organisms are known to be peculiar to certain environments. The presence in ancient rocks in Greenland of fossils of plants which require a warm climate indicates that when those rocks were formed, mild temperatures prevailed in that northerly region. Or, the fact that large numbers of fossils of marine animals are found in rock of era I near the tops of the Appalachians proves that region to have been beneath the sea during that period of time. Or again, the presence of coral, which today is found almost entirely within 20 degrees of the equator, in era I strata in regions which now

have temperate or cold climates, indicates the existence of very widespread mild climatic conditions when those rocks were laid down.

One might continue almost indefinitely to cite examples like those above to show the use that man can make both of rock layers themselves and of the fossils in them in discovering the earth's history. For our purposes, these will serve to illustrate the possibilities involved. But it should be mentioned that our method of approach to the problem carries with it the danger of leaving the reader with an overly simplified picture of the procedure. The geologist working in the field does not take a hurried glance at one or two rock layers and a couple of fossils that they may contain, and then launch into an extensive and dogmatic declaration of the geological history of the region. As a matter of fact, probably no better examples of careful, cautious procedure, in line with the scientific method, can be found in all science than in the field of historical geology. We have cited merely a few specific examples in order to illustrate the procedures and reasoning employed. In actual work conclusions are drawn only after all available clues have been exhaustively studied, and checked and rechecked against each other.

Questions

Discussion

1. Mention the principal features of the planetesimal hypothesis of the origin of the solar system.
2. Describe and criticize the three scientific approaches, mentioned in this chapter, to the determination of the age of the earth.
3. Look up at least one method, other than the ones discussed in this chapter, which has been advanced for calculating the age of the earth.
4. Find out what is meant by the statement, " the half-life of radium is 1690 years."
5. Discuss the significance of conformities and unconformities in determining the geological history of a given region.
6. How would the presence of unconformities affect the reliability of sedimentation as a basis for determining the age of the earth?
7. Mention several clues that glaciers leave which enable one to tell something about their characteristics and behavior.
8. Discuss the limitations of rock correlations as a basis for establishing geological history.
9. Mention the various types of fossils and give examples of each.
10. Suggest a reason why the shorter the period of time during which a

given kind of organism was in existence, the more valuable are its fossils for the determination of geological history.

11. Mention several ways in which plant and animal remains are changed to produce fossils of the " altered remains " type.

12. Discuss the chances that any given plant or animal which is alive today has of becoming a fossil that will be studied by future generations of geologists.

13. Referring back to the preceding question, on an average how do the chances of any given marine animal becoming fossilized compare with those of a land animal?

14. Give reasons for the statement that, " Finding the leg-bone of a horse in the same layer of sedimentary rock with a trilobite fossil would be disastrous to the line of argument employed by geologists in arriving at their conclusions about geological history."

15. Under what sort of circumstances might not the discovery mentioned in the preceding question be disastrous to geological argument?

16. In what way does the geological argument from fossil evidence depend on plants and animals having changed with time?

Multiple Choice

1. Which one of the following is true *and* offers the best reason for considering the radioactive disintegration method the most accurate for determining the age of the earth? (a) It is the most recent. (b) Man has been unable to vary rates of radioactive decomposition. (c) It gives a maximum instead of a minimum figure. (d) It was the first to be used and therefore has been checked over a longer period than any other. (e) It agrees most closely with Archbishop Ussher's figure.

2. Which of the following would offer the best basis for concluding that a certain layer of sandstone in California was deposited at about the same time as one in Florida? (a) Are about the same distance below the surface. (b) Are the same number of rock layers beneath the surface. (c) Have exactly the same chemical composition. (d) Have same physical characteristics such as density, tensile strength, etc. (e) Contain similar fossil remains.

3. An " unconformity " is (a) a fault or crack along which an earthquake has occurred, (b) the plane of division between two layers of lava which failed to fuse together, (c) a fossil discovery which does not agree with other evidence, (d) a vertical plane of cleavage caused by isostatic movements, (e) a certain type of division plane or boundary between adjacent layers of sedimentary rock.

4. Other things being equal, the fossils most valuable in fixing the age of a rock deposit are those of organisms that (a) were the largest of their kind, (b) were in existence for only a short time, (c) existed in the greatest abundance in a given area, (d) changed very little over very long periods of time.

5. In the light of present day evidence, if reliable radioactive calculations place the age of one rock deposit at 1 billion years and of another at 2 billion years, one may conclude that (a) the earth is actually about 1.5 billion years old, (b) the discrepancy proves the method unreliable, (c) the earth is at least 2 billion years old, (d) the earth cannot be more than 1 billion years old.

True–False

1. The vast majority of plants and animals leave no trace in the earth of ever having existed.
2. Bones of large mammals are usually found in rocks of considerable geological age.
3. No fossils of marine animals are ever found on mountain tops.
4. All the scientific methods of calculating the earth's age give minimum figures rather than maximum.
5. The greater the similarity of structure of an ancient and a modern organism, the more valuable are fossil remains of the former.
6. The presence of a conformity between two rock layers indicates no withdrawal of the sea between the times of their being deposited.
7. The fossils of greatest age have usually been altered.
8. Rock correlation methods find their greatest value in dating rock deposits in widely separated regions, with respect to each other.
9. Limestone is never deposited under water.
10. In general marine forms of life stand a better chance of being fossilized than do land forms.

Chapter 10

GEOLOGICAL HISTORY

IN this chapter we shall turn our attention to some of the information that rocks and fossils give us about actual events in earth history. For convenience in discussing the red-letter days in the life of the earth it is desirable to divide its life span of several billion years into smaller intervals. In our consideration of the subject we will discuss and correlate the two bases upon which such a time division is made.

GEOLOGICAL CYCLES

The first of the two schemes is one determined by the earth itself, not one imposed upon it by man. First we shall define some terms in connection with geological cycles and then consider a typical cycle in detail.

Definition of Terms

It will be recalled that in our discussion of " leveling off " versus " building up " factors in Chapter 8 the point was made that at any given time one set or the other might be dominant. The evidence indicates that this position of dominance does not shift back and forth in a haphazard fashion but alternates in a definite periodic manner. The history of the earth's crust has been one of a series of cycles, each of which may be thought of as having had two phases. These are called *revolution* and *submergence* respectively. The revolution stage is a comparatively short period of disturbance on the earth's surface and comes as the climax of an interval during which the building-up agents have been in the ascendancy. The submergence stage is at the opposite extreme and results from the forces of erosion having held the upper hand for a time. It is a period of comparative geological quiet and lasts much longer than does the revolution phase. The in-

terval from one revolution to the next, or, in other words, the period covered by one revolution plus one submergence is called a *major geological cycle*. Such a cycle is, on the average, 150 to 200 million years in length.

Conditions at Maximum Submergence

Maximum submergence occurs when, because of a period of dominance of gradational factors, the leveling-off of the earth's surface has reached its greatest extent. At such a point a large proportion of the earth's surface is covered with water. Few or no mountains are visible, and even hills are largely absent. The regions that remain unsubmerged are reduced to islands, the rest of the continental areas forming the bottoms of shallow seas.

Since there are no steep slopes, mechanical erosion by running water is at a minimum. There is a very considerable amount of chemical erosion in progress, since so much territory is in contact with water. Extensive limestone deposits are typical of a period of submergence.

The combination of widespread seas and a lack of hills and mountains on the exposed land makes for warm, moist, uniform climatic conditions over much of the earth. The reason for this may be illustrated by the manner in which a vessel of water is heated on the stove. In such a case, although the heat is applied at the bottom only, the temperature of the entire amount will rise quite uniformly if the water is able to circulate freely throughout the container. This is due to the convection currents that are set up. The existence of such currents in the atmosphere has already been described in a previous chapter in connection with the discussion of winds. In the same manner, if both water and air are free to circulate on and about the earth, temperatures in all latitudes will be fairly high and quite uniform although, as previously discussed, the sun's heat is being applied principally in the equatorial regions. Such a state of affairs prevails during submergence. At this stage of a geological cycle there are neither desert regions nor areas covered with glaciers. Water, which at other times is tied up in the form of ice around the poles, is released to swell the total volume of the oceans and aid in engulfing the continents.

Obviously, the conditions just described would be ideal for the existence of marine plant and animal life. Consequently, at submergence points in the cycles that have occurred since life began on the earth, such forms have been widespread and abundant. Land life during such an interval naturally is quite limited because of the relatively small amount of dry land. Even here, however, it should be pointed out that as far as climate is concerned, what dry land does exist offers a fairly ideal environment.

Conditions as a Revolution Approaches

The wane of the submergence phase and the approach of the next revolution is heralded by a general raising of the continents. This is accompanied by the beginning of mountain formation.

As slopes become greater, streams flow faster and so mechanical erosion increases. Therefore, one might expect to find the layers of limestone mentioned above overlaid by layers of shale-forming mud and silt, which in turn would be covered with strata of still coarser deposits.

As the rising land begins to interfere with the free circulation of air and water, the climate ceases to be so uniform throughout the world. This decreasing uniformity means that greater extremes of temperature develop in the seas, affecting marine life adversely. As will be brought out in the next section, there is not, however, a corresponding increase and spread in forms of land life.

Conditions at the Climax of a Revolution

At this point in the cycle, the greatest extremes of altitude, temperature, and humidity are found on the earth. The continents are largely exposed with only small areas along the edges being submerged. The earth's surface is marked by great high plains and plateaus, young rugged mountains, and deep valleys containing swiftly flowing streams.

Because of the steep slopes and rapid streams, large amounts of coarse material are carried out and deposited on the ocean beds. The amount of chemical erosion occurring at this time is probably about the same as at the stage of greatest submergence, but, because of the enormous increase in mechanical weathering, it comprises a minimum percentage of the total erosion.

At this time the free circulation of water and atmosphere is most seriously interfered with. Average temperature conditions in some parts of the world are ideal for the accumulation of glaciers, while a tropical climate prevails in other localities. In some areas it may rain only once in several years, while others receive hundreds of inches per year.

The fact that this stage of the cycle finds a maximum amount of dry land exposed, might at first lead one to expect a correspondingly abundant and widespread existence of land life. But such is not the case because, as a result of the temperature and humidity conditions mentioned above, much of the dry land offers far from ideal conditions for land life. What has happened is that during these periods of revolution the most rapid evolution in life forms has occurred. As the environment becomes increasingly severe with the approach of a revolution, types of organisms which have become specialized are faced with the choice either of changing with the environment or of becoming extinct. Specific examples in this connection will be mentioned later in the chapter.

Conditions as a Submergence Approaches

Trends at this stage are essentially the opposite of those discussed for the period leading up to a revolution. Erosion is now dominant over building. Mountains and hills are worn away and the material deposited in the lowlands and on sea floors. Largely because of this deposition of sediment on the continental shelves, sea level is raised and the oceans encroach on the continents. A progressively smaller percentage of total erosion is mechanical in nature. Climatic conditions over the earth increase in uniformity and become more and more favorable for the growth and spread of marine life.

Since man's existence on the earth has been entirely confined to a brief interval near the climax of a revolution, he has never faced the prospect of the decreasing dry land which comes with the approach of a submergence. It is a situation which undoubtedly he will encounter some millions of years from now unless, in the meantime, he manages by his own efforts to eliminate himself from the face of the earth. The effect on human affairs of the approach of this next stage

in the geological cycle provides material for interesting speculation. In the first place, oceans would not have to rise a great deal to engulf a considerable area of what is now dry land. An increase in sea level of 500 feet, which is just slightly more than 1 per cent of the oceans' present maximum depth, would bring the Gulf of Mexico up to Cincinnati, while a 1,000-foot rise would take it to the arctic regions. With such a rise all land life would be forced to confine itself to a smaller and smaller area. One can imagine the wars that would be fought, always with the noblest of expressed motives, over this steadily dwindling bit of dry land. Yet, what to man would be a major catastrophe, from the standpoint of the earth would be but the natural recurrence of a particular stage in the normal geological cycle, a stage that is reached merely because, for a time, leveling becomes dominant over building. It would no more represent a calamity to the earth than does the seasonal drying up of a pond which results in wiping out its particular frog or fish population.

In the above paragraphs we have traced the characteristics of various stages of the geological cycle. There is well-established fossil evidence for the occurrence of three such cycles covering the last half billion or so years and every reason to suppose that these three were preceded by others of like nature. It should be remembered that in these major cycles the trends have been world-wide. Superposed upon these there have been many minor cycles, similar in nature but less extreme and more localized. In other words, the earth does not move smoothly and uniformly from one stage to the next of a major cycle but rather in a slow spasmodic fashion through a series of minor disturbances.

Now we will turn our attention to a second basis upon which earth history is subdivided, after which we will fit specific events into the picture.

Geological Eras

The second scheme of division is a man-made one which divides earth time into intervals called eras. These in turn are further divided into smaller units known as periods.

Names of the Eras

The general pattern in terms of which the eras are given names is somewhat analogous to that which serves as a basis for the division of human history. What may be thought of as well-recorded human history begins with the appearance of peoples who left written records. Similarly, well-recorded earth history may be defined as beginning with the appearance on the earth of plants and animals which left fairly complete and well-connected fossil records. We divide well-recorded human history into three parts, namely ancient, medieval and modern. Of these, ancient history covers the greatest extent of time, medieval the next greatest, and modern the least. In an analogous fashion, well-recorded earth history is separated into three intervals called eras. From the stage of development of life at the time, as evidenced by the fossil remains of plants and animals, these are name *Paleozoic* or era of old life, *Mesozoic* or era of intermediate life, and *Cenozoic* or era of recent life. Here also, the most recent covers the least time and so on back through the three. These are the eras that were designated as I, II, and III respectively in the previous chapter.

Then in human history there is a period of time earlier than that for which we have written records. This interval, which exceeds historic time in length, we call prehistoric time. During it there was some human life in existence and some scattered records are available but we do not have a well-connected world-wide story of human development throughout these prehistoric ages. Likewise in earth history there was a time preceding the Paleozoic era when life existed on the earth but for which fossil records either are missing entirely or are poorly preserved and unconnected. These geological prehistoric times are divided into two eras which, in chronological order, are called Archeozoic or era of earliest life, and Proterozoic or era of earlier life.

We have no exact knowledge of events during earth time before the beginning of life. Authorities divide it into either one or two eras. Our purposes will be served adequately if we consider it all to belong to the Azoic era or era of no life.



FIG. 61. A view in the Grand Canyon of Arizona showing relations of Archeozoic, Proterozoic, and Paleozoic rocks to each other. V, Archeozoic metamorphoses; B, H, and Sh, Proterozoic partly metamorphosed strata with associated lava, d; T, BA, and M, Cambrian strata; R, Mississippian and Ss, Ssh, C, and K, later Paleozoic — Au, unconformity between Proterozoic and Paleozoic. A vertical mile of strata is in view. Photo by Carkhuff. (After Noble, U. S. Geol. Survey Bulletin 549.)

The boundaries between eras are determined in general by major revolutions. As was mentioned previously, eras are subdivided into periods largely on the basis of the occurrence of minor geological disturbances.

Principal Events

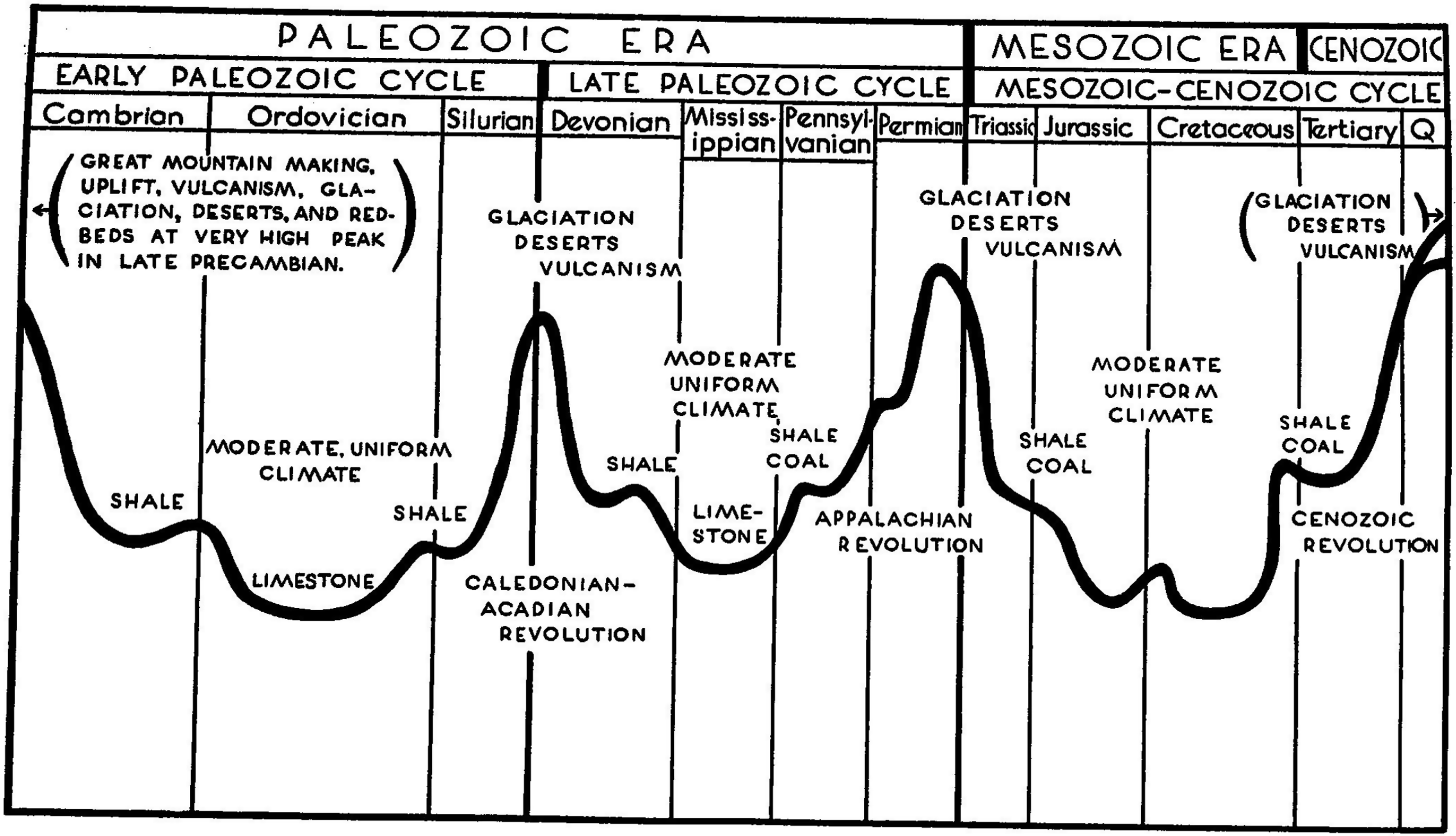
Since a detailed description of even the outstanding events of geological history would make much too long a story for inclusion in this book, this part of the discussion has been largely put into table form. Table 5 will give the reader an idea of the relative lengths of the various eras. The extent of each one, as shown in the second column, is an estimate based on an assumed total geological age of three billion years. The last column gives the lengths of the eras as they would be on a scale model for which all earth time is set equal to one calendar year. Points to be particularly noted are that more than half of the earth's present life had elapsed before the beginning of what we have called well-recorded earth history, that each era was longer than the one that followed it, and that the present one, the Cenozoic, has extended over only a tiny fraction of the total.

TABLE 5.[1] LENGTHS OF ERAS ASSUMING EARTH TO BE THREE BILLION YEARS OLD

Era	*Length in Millions of Years*	*Extent on Scale of All Earth-Time = 1 Year*
Azoic	1000	Jan. 1 to May 13
Archeozoic	800	May 14 to Aug. 13
Proterozoic	650	Aug. 14 to Oct. 28
Paleozoic	350	Oct. 29 to Dec. 8
Mesozoic	140	Dec. 9 to Dec. 25
Cenozoic	60	Dec. 26 to Dec. 31

[1] After table in *Down to Earth*, Croneis and Krumbein, p. 285. Chicago: Univ. Press, 1936.

Next let us call the reader's attention to Fig. 62 in which the three major geological cycles of well-recorded earth history are related graphically to the eras and periods which extend over the same interval of time. It will be noted that the Paleozoic era included two


PALEOZOIC ERA
MESOZOIC ERA
CENOZOIC
EARLY PALEOZOIC CYCLE
LATE PALEOZOIC CYCLE
MESOZOIC-CENOZOIC CYCLE
Cambrian
Ordovician
Silurian
Devonian
Mississippian
Pennsylvanian
Permian
Triassic
Jurassic
Cretaceous
Tertiary
Q
GREAT MOUNTAIN MAKING, UPLIFT, VULCANISM, GLACIATION, DESERTS, AND RED-BEDS AT VERY HIGH PEAK IN LATE PRECAMBIAN.
SHALE
MODERATE, UNIFORM CLIMATE
LIMESTONE
SHALE
GLACIATION
DESERTS
VULCANISM
CALEDONIAN-ACADIAN REVOLUTION
MODERATE UNIFORM CLIMATE
SHALE
LIME-STONE
SHALE COAL
APPALACHIAN REVOLUTION
GLACIATION
DESERTS
VULCANISM
SHALE COAL
MODERATE UNIFORM CLIMATE
GLACIATION
DESERTS
VULCANISM
SHALE COAL
CENOZOIC REVOLUTION

of the major cycles while the third one extended over the two most recent eras.

Now, with the relationships of Table 5 and Fig. 62 in mind, the reader is in a position to study Table 6. Here the eras and periods are listed in chronological order and for each one there are given some

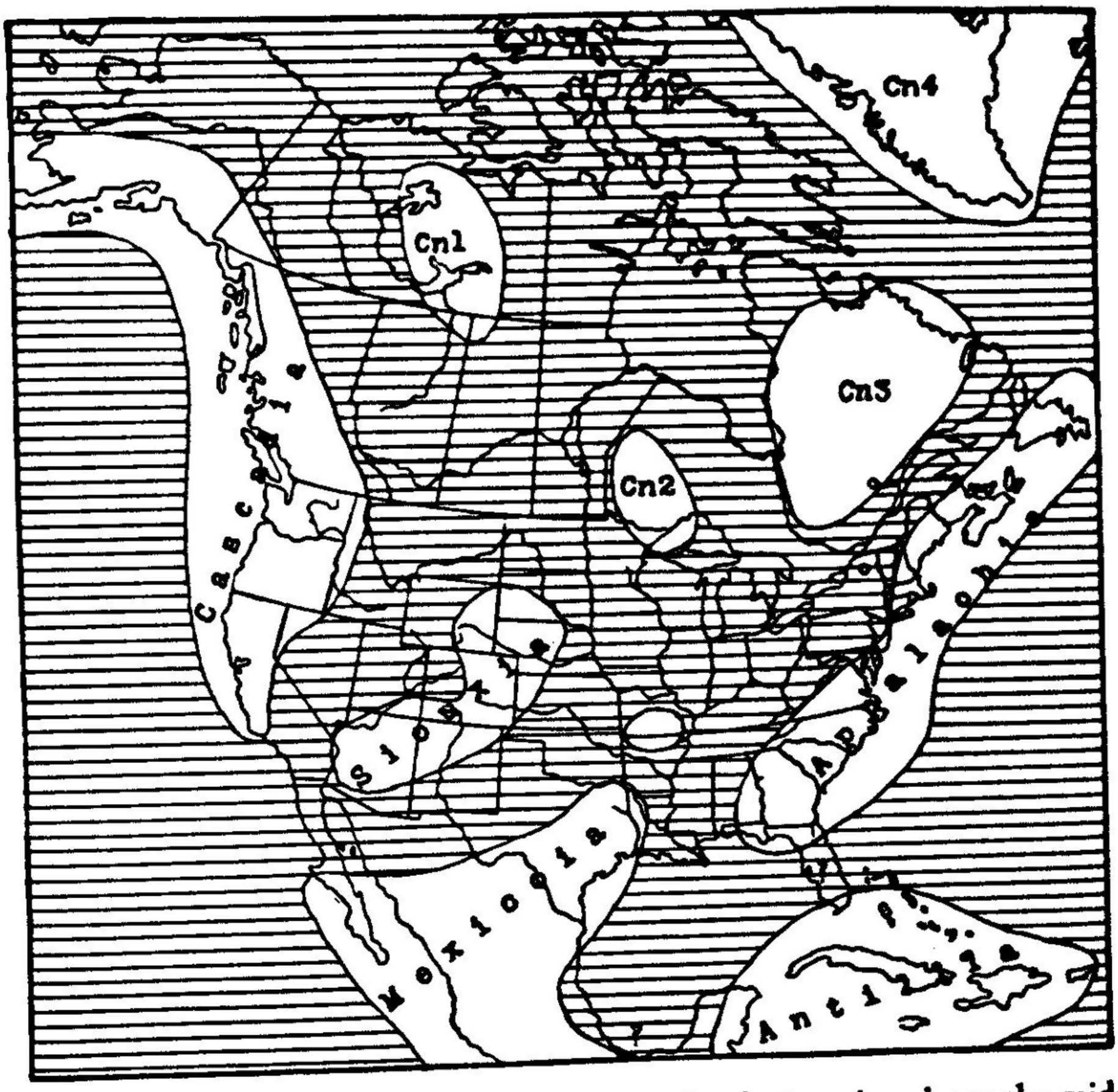


FIG. 63. Map showing sea and land areas in North America about the middle of the first Paleozoic cycle. White areas are land, shaded ones, water. (From maps by B. Willis, C. Schuchert, and R. Chamberlin.)

of the main geological events, the principal economic products, and the chief forms of life. It will be noted that the geological events mentioned are almost exclusively those occurring on the North American continent since it would be impossible in the space available to tell a world-wide story.

TABLE 6. A TABLE WHICH LISTS BRIEFLY THE GEOLOGICAL DESCRIPTION AND CONDITIONS OF LIFE FOR EACH OF THE ERAS AND PERIODS OF EARTH HISTORY

Era	Period	Description	Life
Azoic 1,000 M. Y.		An era by definition. Assumed to cover period from earth's origin to beginning of life. Events would include acquisition by earth of atmosphere and hydrosphere.	None
Archeozoic 800 M. Y.		Archeozoic rock underlies almost all other strata. Largely metamorphic and igneous. Types of former indicate sediments of era much the same as those of today. Latter types indicate great volcanic activity. Examples of Archeozoic rock include Laurentian Highlands, core of Rockies, and parts of Ozarks and Black Hills. Era ended with raising up of Canadian Shield, called the Laurentian Revolution.	Fossil remains confined to a few sponges and blue-green algae. Presence of large limestone and graphite beds taken as presumptive evidence of life since living organisms involved in making such deposits. Conclusions regarding events based largely on rock criteria. Dearth of fossils does not indicate paucity of life since metamorphism would destroy most fossil remains.
Proterozoic 650 M. Y.		Rock still predominantly metamorphic although some sediment. Two periods of glaciation. During first, regions as far apart as North America, Australia, and Africa covered with ice sheets. Valuable mineral deposits include Lake Superior iron ore, Michigan copper, and Canadian gold, silver, and nickel.	Direct evidence confined to a few fossil remains of one-celled plants, sponges, and worms. As in rock of the preceding era, large deposits of limestone and graphite imply the existence of various forms of plant and animal life.

Era	Period	Description	Life
Paleozoic 350 M. Y.	Cambrian	At beginning of period large mountain ranges along eastern and western coasts of North America, called Appalachia and Cascadia respectively. Only submerged portions of continent were trough-like depressions along inland edges of these mountains. By the close of the period about a third of continent submerged. Rock of period contains gold in Black Hills and lead in Missouri.	Marks beginning of well recorded earth history. Fossil evidence indicates abundant life. Many forms of invertebrates including trilobites and brachiopods. No vertebrate fossils ever found in Cambrian rock.
	Ordovician	By middle of period about 60% of North America submerged (See Fig. 63). Taconic Mountains formed. Products of Ordovician rock include oil and gas in Ohio, Indiana, Kentucky and Oklahoma; also lead and zinc in Illinois, Wisconsin, and Iowa.	Increasing abundance and complexity of invertebrate life. Appearance of first vertebrates. Only 10 or 15 of the thousand or more forms of Cambrian life survived the minor revolution to become part of the several thousand forms known in the Ordovician.
	Silurian	Most of North America covered by shallow seas which extended from Gulf of Mexico to Arctic Ocean and from New York to Iowa. Major disturbance at close of period marked end of first Paleozoic major cycle. Important products include salt and gypsum in Ohio and New York, and Appalachian iron ore.	First appearance of true fishes, scorpions, and air breathers First coral reefs formed.
	Devonian	Close of period marked by raising up of Acadian Mountains. Chief North American product of Devonian rock is limestone used in Portland cement.	Dominant forms of life were fishes, among which was a primitive type of shark. Bony fish small and few in number. Appearance of first amphibians. First extensive land vegetation.

Era	Period	Description	Life
	Mississippian	Greatest submergence of North America between Ordovician and present. Raising up of Appalachians begun toward close of period. Products are Bedford limestone of Indiana and Berea sandstone of Ohio.	No outstanding innovations in life forms. Widespread warm seas offered ideal environment for marine plants and animals.
Paleozoic (continued) 350 M. Y.	Pennsylvanian	During most of period Canada was above water while a large shallow sea covered western U. S. Some 90% of all coal mined in North America today comes from rock of this period, including almost all mined east of Kansas. Other Pennsylvanian products include large petroleum deposits in Oklahoma and Kansas.	Amphibians dominant form of animal life. Dominant plants were huge club mosses and their relatives which constituted the vegetation from which today's most valuable coal deposits came. Oldest fossils of reptiles found in Pennsylvanian rock.
	Permian	Permian was a period of mountain building throughout the world. Also included greatest glaciation world has known. Appalachians were completed by close of period. Among important products are large Texas petroleum deposits, and phosphate in Wyoming, Montana, Utah, and Idaho.	Being a period of revolution, Permian was marked by many changes in life forms. Vertebrate land life consisted of amphibians and reptiles which had evolved into entirely different species by beginning of next period. Similar transformations occurred in land plants.
Mesozoic 140 M. Y.	Triassic	With this period earth entered upon third major cycle of well recorded earth history. North America almost entirely exposed at beginning, but Pacific Ocean began to encroach toward period's close. Scarcely any mineral products of value in Triassic rock.	Period marked by rise of reptiles as evidenced by appearance of dinosaurs as well as of forms which could swim and fly. Mammals appeared but were small and unimportant.

Era	Period	Description	Life
Mesozoic (continued) 140 M. Y.	Jurassic	Jurassic seas covered much of western U. S. Eastern part of country exposed throughout period. Oceans gradually receded toward close of period, the end being marked in the U. S. by the raising up of Sierra Nevada Mountains. Rock contains some important gold deposits in California and coal in Alaska.	First birds appeared. They were distinctly reptile-like in appearance and evidently evolved from that form. Reptiles continued to advance and attained dominance so that Mesozoic era often called the "Age of Reptiles."
	Cretaceous	Period marked in North America by repeated advances and recessions of oceans. At maximum submergence a 1,000-mile-wide sea joined Arctic Ocean and Gulf of Mexico. Revolution which began toward close of period is not yet finished. Rocky Mountains raised up at this time. Principal products of Cretaceous rock are coal and oil, the latter from Texas and Mexico. About one-sixth of all petroleum produced thus far comes from rock of this period.	Reptiles reached their greatest development. Climax came with appearance of huge dinosaurs, some of which reached lengths of 100 feet and weighed 30 tons (see Fig. 64). By close of period flowering plants were widespread as also were many types of trees. The latter were similar in form to modern trees but of different species.
Cenozoic 60 M. Y.	Tertiary	Tertiary period taken to include all but last million years of Cenozoic era. Maximum submergence early in period when Gulf extended north to Ohio River. At close of period North America was much as now. Andes, Alps, Himalayas, and Pyrenees all raised up in Tertiary. About half of all petroleum produced in world comes from rock of this period. Other valuable products are copper, silver, and lead.	Rapid evolution of new forms of plants and animals. Dinosaurs entirely disappeared because of inability to adapt themselves to changing conditions. Reptiles as a whole became of minor importance and mammals acquired dominance. Most mammals of this period belonged to species no longer in existence. Fossils of primitive primates found in Tertiary rocks.

Era	*Period*	*Description*	*Life*
Cenozoic (continued) 60 M. Y.	Quaternary	Quaternary began with advent of interval of widespread glaciation. Evidences of advance and recession of at least four ice sheets over parts of North America, the last one withdrawing some 20,000 years ago. Interval since this last ice age known geologically as "the present." Impossible to be certain whether ice ages are over with until next revolution, or whether earth is now simply between two of them.	So-called "Age of Man" began with Quaternary. Earliest man of period belonged to species known as Dawn Man. Modern man, or Homo sapiens, appeared about close of last ice age. This brings Neolithic man into the picture at 11:57 P.M. on December 31 of the one-year scale model described in Table 5.

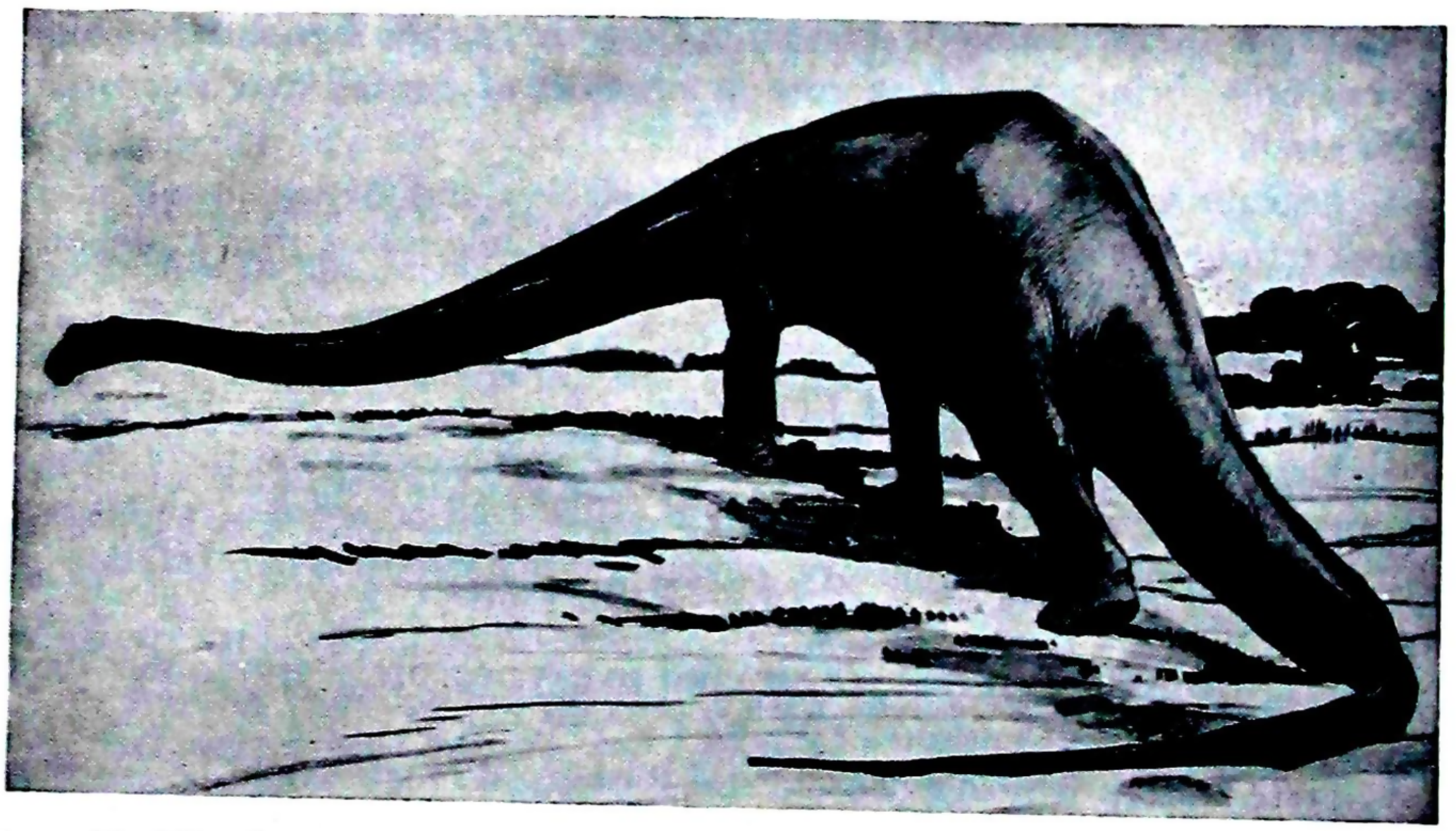


FIG. 64. The longest of all known dinosaurs, a sauropod, Diplodocus. A mounted skeleton in the Carnegie Museum of Pittsburgh measures 87 feet long. Restored by C. R. Knight under the direction of H. F. Osborn. (Courtesy American Museum of Natural History.)

This very hurried tour through geological history with brief side excursions into the fields of life and economic products probably should not be concluded without some mention of the earth's future. Naturally, predictions about the geological future can be made only in terms of the past. On such a basis it seems fairly safe to prophesy that, barring some great cosmic catastrophe such as collision with an-

other heavenly body or similar occurrence, our planet will continue to move calmly and inexorably through one geological cycle after another, quite undisturbed and unaffected by the antics and activities of man and the other living creatures which live their brief life spans upon its ever-changing surface.

Questions

Discussion

1. Define the terms " revolution " and " submergence " as applied to geological cycles and relate them to the two types of agents discussed in Chapter 8.

2. List the principal characteristics of earth contours, climate, and life to be expected at each of the following points in a major geological cycle: (a) At maximum submergence; (b) At climax of a revolution; (c) Halfway from a submergence to a revolution.

3. Relate the discussion of weather in Chapter 7 to the fact that approaching revolution brings with it increasing extremes of rainfall over the surface of the earth.

4. Explain why it is that the greatest glaciers and largest deserts occur on the earth at the same point in the geological cycle rather than at opposite phases.

5. Find out how close to your locality the oceans would be brought by a rise in sea level of 500 feet; of 1,000 feet.

6. Describe the analogy drawn in this chapter between the three periods of human history and the three eras of so-called well-recorded earth history.

7. Well-recorded earth history covers approximately what fraction of total earth-time?

8. Name three events that are assumed to have occurred during the Azoic era.

9. State reasons why Archeozoic and Proterozoic rock would be expected to be much more largely metamorphic in nature than Mesozoic rock.

10. During what era did the first widespread glacial age for which we have good evidence occur and how extensive was the glaciation?

11. Name the eras indicated in each of the following:

a. Era in which man made his earliest appearance.
b. Era during which the uplift of the Rocky Mountains began.
c. Era that corresponds to medieval human history.
d. Era during which amphibians reached their greatest development.
e. An era during which the Gulf of Mexico and the Arctic Ocean were joined by a large inland sea.

12. Identify the three eras of well-recorded earth history with the last three major geological cycles.

13. When did the Caledonian revolution occur?

14. Compare the life in the last cycle of the Paleozoic era with that of the Mesozoic era with regard to types and abundance.

15. On what sort of a basis are eras divided into periods?

16. Identify as to era and period, the rocks in which the following are found: (a) Lake Superior region iron ore; (b) Pennsylvania coal; (c) Pennsylvania oil; (d) California gold.

17. Give at least two reasons why no man ever used a domesticated dinosaur as a beast of burden.

18. In what era did all of the following occur: mammals made their appearance; the earliest birds came upon the scene; flowering plants began.

19. What are the names and the relative lengths of the two periods into which the Cenozoic era is divided?

20. The earth is believed to be at what point in the geological cycle at the present time? Give reasons.

21. What is the principal North American economic product of Cenozoic rock?

22. Find out all you can about the geological history of your own locality.

Multiple Choice

1. In the analogy between human and earth history, ancient in the former corresponds to the — (a) Proterozoic, (b) Archeozoic, (c) Cenozoic, (d) Paleozoic, (e) Mesozoic — era.

2. List all of the following that would be true at the climax of a major geological revolution. (a) Minimum distribution of marine life over the surface of the earth. (b) Greatest extremes of rainfall over the earth. (c) Most favorable conditions for both glaciers and deserts to exist on the earth at the same time. (d) Maximum continental area out of the sea. (e) Comparatively rapid evolution of plant and animal forms.

3. During which era were the most valuable coal deposits laid down? (a) Archeozoic. (b) Azoic. (c) Cenozoic. (d) Mesozoic. (e) Paleozoic. (f) Proterozoic.

4. Which one of the above eras is known as the " age of reptiles "?

5. If the entire age of the earth is set equal to 1 year, man has been in existence (a) about 10 seconds, (b) between 1 and 5 minutes, (c) almost 2 weeks, (d) about 1½ days, (e) 30 seconds.

True–False

1. The last glacial age in North America is believed to have ended less than 40,000 years ago.

2. Paleolithic man lived during the Paleozoic era.

3. Erosional processes proceed much more rapidly at maximum revolution than at the climax of a submergence.

4. Greatest uniformity of climate over the earth and most rapid evolution of life occur at the same stage of the geological cycle.

5. The last three geological cycles have included three geological eras.

6. There is presumptive evidence of the existence of life at a time earlier than that at which fossil records begin.

7. Geological eras are subdivided into periods.

8. More than half of the earth's total history is without well-connected fossil records.

9. The greatest submergence of North America occurred during the Paleozoic era.

10. Dinosaurs are believed to have disappeared chiefly because of the attacks made on them by large carnivorous mammals.

Chapter 11

NUMBERS AND NUMBER SYSTEMS

Introduction

In his travels along the road of scientific progress, man has discovered from time to time that further advance required the development of additional tools. Among the most important such tools have been numbers, number systems, and the various branches of mathematics. In a book of this sort it would be neither possible nor desirable to treat all of these in detail, but we will try to see something of the origin and development of numbers and number systems, and investigate certain of the concepts involved in their manipulation.

In view of the fact that the very mention of anything even faintly suggestive of mathematics seems to throw so many people into a kind of mental panic, let us begin this discussion by seeing how an outstanding scientist of the present day looks upon this bugaboo. In his book, *Mathematics for the Millions*, Lancelot Hogben presents a concept of the term which is well worth our consideration. He suggests that mathematics should be thought of simply as a language of *size* in contrast to our ordinary, everyday language, which is one of *sort* or *kind*. This language of size provides a medium for answering such questions as "How much?" or "How many?" whereas the language with which we are more familiar makes it possible to answer such queries as, "What kind?" or "What sort?" In discussing this point, Hogben [1] tells the following story about Diderot, the French encyclopedist, and Euler, the most famous mathematician of his day:

Diderot was staying at the Russian court, where his elegant flippancy was entertaining the nobility. Fearing that the faith of her retainers was at stake, the Tsarita commissioned Euler to debate with Diderot in public.

[1] Hogben, *Mathematics for the Millions*, p. 13. New York: Norton & Co., 1937.

Diderot was informed that a mathematician had established a proof of the existence of God. He was summoned to the court without being told the name of his opponent. Before the assembled court, Euler accosted him with the following pronouncement, which was uttered with due gravity:

$$\text{“}\frac{a + b^n}{n} = x\text{; donc Dieu existe — répondez! ”}$$

Algebra, however, was Arabic to Diderot. Unfortunately he did not realize what the trouble was. Had he realized that algebra is just a language in which we describe the sizes of things in contrast to the ordinary language which we use to describe the sorts of things in the world, he would have asked Euler to translate the first half of the sentence into French. Translated freely into English, it may be rendered, “ A number x can be got by first adding a number a to a number b multiplied by itself a certain number of times, and then dividing the whole by the number of b's multiplied together. So God exists after all. What have you got to say now? ” . . . Euler's troubles would have begun when the court wanted to know how the second part of the sentence follows from the first part. Like many of us, Diderot had stage-fright when confronted with a sentence in size language. He left the court abruptly amid the titters of the assembly, confined himself to his chambers, demanded a safe conduct, and promptly returned to France.

Our modern world is so organized that most of us are considerably more familiar with size language than was poor Diderot. At least, most of us have learned to talk it fairly well, although we may write it somewhat less readily. Our everyday life is so concerned with numbers that we have to be able to speak the language of size with some facility. To quote Mr. Hogben [2] again,

> The first men who dwelt in cities were talking animals. The man of the machine age is a calculating animal. We live today in a welter of figures: cookery recipes, railway time tables, unemployment aggregates, fines, taxes, war debts, overtime schedules, speed limits, bowling averages, betting odds, billiard scores, calories, babies' weights, clinical temperatures, rainfall, hours of sunshine, motoring records, power indices, gas meter readings, bank rates, freight rates, death rates, discount, interest, lotteries, wave-lengths, and tire pressures. Every night, when he winds up his watch, the modern man adjusts a scientific instrument of a precision and delicacy unimaginable to the most cunning artificers of Alexandria in its prime. So much is commonplace. What escapes our notice is that in doing these things we have learned to use devices which presented tremendous difficulties to the most brilliant mathematicians of antiquity. Ratios, limits, accelerations are not remote abstractions dimly apprehended by the solitary genius. They are photographed upon

[2] *Ibid.*, p. 16.

every page of our existence. . . . We constantly find that we have no difficulty in answering questions which tortured the minds of very clever mathematicians in ancient times. This is not because you and I are very clever people. It is because we inherit a social culture which has suffered the impact of material forces foreign to the intellectual life of the ancient world. The most brilliant intellect is a prisoner within its own social inheritance.

The above quotation concerns principally the relation of this size language to our everyday lives. As might be supposed, it is of even greater and more fundamental importance when we confine the discussion to science. The reader will recall that science was defined in Chapter 1 as " that body of knowledge which has been obtained by controlled, quantitative investigation." The very wording of this definition suggests that a body of knowledge is scientific just to the extent that its results can be expressed in numerical form. To express results numerically means that the quantities involved must be measured, and these both imply the use of a language of size. The matter of measurement and units of measurement will be considered in detail in the next chapter. Now we will proceed to the main business of this chapter, as previously mentioned.

Number Names and Bases

From Counting to Number Names

Man's use of numbers may be traced back to his first attempts to count things. His first use of counting probably was due to two needs, both connected with agricultural pursuits. The first concerned the matter of keeping track of his flocks and herds. It was necessary to have some means of knowing how many animals there were, and what the increases and decreases amounted to. As long as the flocks were small, he gave names to the individual animals and checked them by going over the list, a process of the sort we follow today in " human flocks " when we call the roll. But this method was scarcely practical with large herds, and counting provided a solution to the problem. The other need was that of counting time. When man began to plant grain and domesticate animals which bear their young at certain times of the year, he needed to be able to keep track of the seasons with a fair degree of accuracy.

For many generations after he began to count, man gave no names

to his numbers but used what may be called " counters." The most important of these were his fingers, whose use in such a connection has not entirely disappeared today as evidenced by bidding procedures on the grain exchange and by the public speaker who thus counts off his " firstly," " secondly," and " thirdly." In addition to fingers and toes, other objects such as pebbles and sticks were used. Later on man began to attach names to his numbers. Building up a complete set of number names is a very complicated and difficult job and progress was slow. As a matter of fact, there are tribes in existence today who have no names for numbers greater than three or four. For larger quantities they simply use a word which means several or many.

It is interesting to note that primitive man's number names invariably referred to some specific object or application directly connected with that number. The Chinese word for " two " is the same as the word for ears. Other peoples have used words which mean " eyes " or " wings " to signify " two." Certain Indian tribes in Paraguay use for " four " a word whose literal translation is " ostrich toes." The Zulus have a word for " six " which means " taking the thumb," that is, a full hand of five fingers plus the thumb from the other hand. For a long time man was quite unable to think of numbers in the abstract. To reach that advanced state of mind he had to make the same transition that every child makes in school when he progresses from an understanding of the concrete

$$5 \text{ oranges} + 2 \text{ oranges} = 7 \text{ oranges},$$

to an appreciation of the abstract,

$$5 + 2 = 7.$$

The latter statement marks a considerable advance in thinking since it is a mathematical fact that can apply not only to oranges but also to stars, or dollars, or pounds of butter, or anything else.

Number System Bases

Our number system, which is of the decimal variety, makes use of a base of 10. We count up to 10 and then start over again with 10 plus 1, or 11; 10 plus 2, or 12; and so forth. This idea of counting things in groups seems to have been hit upon by man rather early in his mathematical history, and all sorts of number bases have been

employed. It will be obvious that for any given scale or base, there must be at least as many symbols as there are numbers in the base. In our scale of 10 we use the 0 and the digits 1, 2, 3, 4, 5, 6, 7, 8, and 9. From 9 on we repeat, using zero as a place-holding symbol. Since man's fingers and toes figured so importantly in his early counting procedures it was quite natural that they should become bases of number scales. Consequently, almost all number systems that have ever been used have employed five or some multiple of five as a base, with ten being the most popular.

Some early peoples counted in twenties. Vestiges of this practice remain today in the use by the French of such expressions as " quatre vingt " or " four twenties " instead of eighty. The frequent Biblical references to score in such expressions as " three score and ten " also indicate this use of a base of 20. The early Babylonians counted with a system that had 60 as its base, which meant that they had to have separate and distinct number names and symbols for all digits up to 60.

Some of our common relations such as 12 units = 1 dozen, 12 inches = 1 foot, 12 ounces = 1 pound troy, indicate that a base of 12 has also had some use in the past. In fact, such a *duodecimal* system has some definite advantages over a decimal one. For one thing, fractions are more easily manipulated in such a scheme because of the fact that 12 is evenly divisible by 1, 2, 3, 4, and 6, whereas only 1, 2, and 5 can be divided an exact, whole number of times into 10.

The Writing of Numbers

The Beginning of Number Writing

When man reached such a state of advancement that he wanted to write and record numbers as well as talk about them, a whole new set of difficulties entered the picture. For a while he was able to get along by combining number writing with picture writing, indicating a certain number of objects by simply drawing pictures of that many. The American Indian who wished to record the fact that three men took a six-day canoe trip, drew a picture which showed a canoe with three men in it and alongside, six pictures of the sun. The mathematical limitations of such a number system are obvious.

It is probable that the first actual numbers used by man were simply marks such as notches in sticks, or lines drawn on stone, the number of notches or lines indicating the desired numeral. So the writing of numbers began apparently with a procedure of the sort which we today call "tallying." Some of the problems that arose in going from such a system to our present-day methods can be best discussed by considering a few specific examples.

The Roman Number System

In the Roman system of numerals, which is more or less familiar to most people, the use of the symbols I, II, and III seems to bear out the previous suggestion with regard to tallying. There is some evidence that the use of V for five resulted from the practice of drawing a slanting line across four vertical ones. X, or ten, may have come from combining two V's, one inverted. The use of IIII for four was replaced by IV when the latter proved less difficult to recognize. In IV and also in IX, we see the operation of a sort of principle of position, in that the I has a little different significance when it appears to the left of V or X than when on the right. We shall see a little later that a much more valuable use of this idea of giving significance to the position of a symbol entered the picture when zero was invented.

An important shortcoming of Roman numerals as well as of many other early systems was that many more symbols were required than those for the first ten numbers. Thus XXXXX proved too cumbersome as a symbol for 50 and L was adopted. Then came C for 100, D for 500, M for 1,000. In other words, as larger and larger numbers came into use, new symbols were constantly being required. It should be noted that the Roman was a decimal system.

The Greek Number System

The Greeks also used a decimal system. They had, however, two separate methods of writing numbers. The earlier of these employed symbols which were connected with the initial letters of the number names. About 500 B.C. there came into use a notation in which the letters of the Greek alphabet also served to represent numbers. The table given below will make clear the manner in which the numbers were assigned to letters. In order to account for all the necessary numbers it was necessary to add three letters from an older Greek

alphabet. Two of these appear in the table as symbols for 90 and 900. It will be noted that lower case letters were used and followed by a short line to indicate that they represented numbers rather than letters. A stroke placed before a symbol meant that it was to have 1,000 times its normal value, while the combination of an M with it multiplied it by 10,000.

TABLE 7.[1] SAMPLE NUMBERS WRITTEN IN GREEK NOTATION.

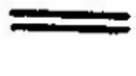
$\alpha', \beta', \gamma' \ldots \iota'$	= 1, 2, 3 . . . 9, 10
$\iota\alpha', \iota\beta', \iota\gamma' \ldots \iota\theta'$	= 11, 12, 13 . . . 19
$\kappa', \lambda', \mu' \ldots \pi', \varphi'$	= 20, 30, 40 . . . 80, 90
$\kappa\alpha', \kappa\beta', \lambda\alpha', \lambda\beta'$, etc.	= 21, 22, 31, 32, etc.
$\rho', \sigma', \tau' \ldots \omega', \text{͡}\varphi'$	= 100, 200, 300 . . . 800, 900
$\rho\iota\alpha', \rho\kappa\gamma', \omega\mu\beta'$, etc.	= 111, 123, 842, etc.
$/\alpha\ /\beta\ /\gamma$, etc.	= 1000, 2000, 3000, etc.
$\overset{\beta}{\mathrm{M}}, \overset{\gamma}{\mathrm{M}}$, etc.	= 10,000, 20,000, 30,000, etc.

[1] American Council of Education, *The Story of Numbers*, p. 15.

The Mayan Number System

Both the Roman and Greek, as well as most other ancient systems, lacked the one element which makes it possible to write numbers of any and all sizes with only as many symbols as the value of the base of the system. That element is the zero, and it is one of the most important features of our modern way of writing numbers. The earliest use of zero, as well as of the principle of position, probably occurred in the number system developed by the Mayas of Central America and Southern Mexico. The system they employed was a vigesimal one, that is, it had a base of 20 and so it was necessary to have symbols for all numbers from one to 19. These were made by combining dots and bars, the former counting one each and the latter five. The symbol for zero is shown below. The following table shows a typical group of these numbers, all less than 20.

TABLE 8. A GROUP OF NUMBERS SMALLER THAN 20, IN MAYAN NOTATION.

0 =	7 =
1 =	10 = 
2 =	11 =
4 =	15 = 
5 =	19 =

For numbers larger than 19, a principle of position came into operation. We make use of this idea in our number system by having what we call units place, tens place, hundreds place, and so on. We understand that to place a number symbol in the tens column multiplies its face value by 10, or to put it in the hundreds column multiplies it by 10 times 10, or 100. The Mayan application of this principle was different from ours in two main respects. In the first place, they wrote their numbers vertically instead of horizontally with the lowest position corresponding to our units place. Then, since the system was based on twenty instead of ten, a symbol written in the second position from the bottom had 20 times its value in the lower place. For reasons connected with Mayan use of an official year of 20×18 or 360 days, placing a number in the third position from the bottom multiplied its value by 20×18, instead of by 20×20 as would have been the case if the system had been entirely analagous to ours. The fourth position gave a value of $20 \times 18 \times 20$ to the symbol. The highest number found in the various inscriptions of the Mayas that have been studied is one of more than 12 million. A few illustrations of numbers larger than 19 written in the Mayan notation are given in Table 9.

TABLE 9. ANALYSIS OF A GROUP OF LARGER NUMBERS WRITTEN IN MAYAN SYSTEM.

Arabic	*Mayan*	*Lowest Position*	*2nd Position*	*3rd Position*
20 =	 =	0	+ (1 × 20)	
79 =	=	19	+ (3 × 20)	
193 =	=	13	+ (9 × 20)	
360 =	=	0	+ (0 × 20)	+ (1 × 20 × 18)
2176 =	=	16	+ (0 × 20)	+ (6 × 20 × 18)

One more table may be of interest in comparing the various systems of writing numbers. Table 10 lists a series of numbers which in our system involves *only two* different symbols. The same numbers are shown as they would be written in the other three schemes that have been considered.

3. List as many of your activities on an average day as you can and check all those which in any way involve numbers or measurement.

4. Trace the development of numbers from counting to the adoption of names for specific numbers.

5. In a number system with a base of eight, what might be the names of the numbers which, in the decimal system, we call four, nine, sixteen, and one hundred?

6. Answer the above question substituting a base of five.

7. List a few advantages that a number system with a base of twelve might have over our present system.

8. Suggest a reason for the fact that most number systems have used some multiple of five as a base.

9. Mention several reasons for believing that number bases of twenty and twelve have had some use in the past.

10. Discuss the importance of a symbol for zero in a number system.

11. Mention five numbers that can be written with fewer symbols in the Roman numeral system than in ours.

12. Mention several ways in which the Greek number system differed from the Roman.

13. Write your age, your weight, and your house number in Roman, Arabic, and Mayan numerals.

14. Trace the principal points in the development and spread of the Hindu-Arabic system of numerals.

15. Write 1492, 1066, 23, and 739 as Roman numerals and then add them using the system of Roman numeral addition described in this chapter.

16. Write 345 and 7 as Roman numbers and then multiply them together according to the method described in this chapter.

Multiple Choice

1. The invention of zero (a) occurred very early in mathematical history, (b) would have been impossible in any but a decimal system of numbers, (c) is generally attributed to Galileo, (d) increased the usefulness of the other digits for writing large numbers, (e) occurred first in the Roman numeral system.

2. We would write the Roman number MLXVI as (a) 1492, (b) 1066, (c) 1606, (d) 1776, (e) 1516.

3. In a number system with a base of eight instead of ten, which one of the following would represent the number that we call 9? (a) 7, (b) 10, (c) 11, (d) 17, (e) 21.

4. Which of the following systems (one or more) have or had a zero? (a) Hindu-Arabic, (b) Mayan, (c) Roman, (d) Greek.

5. In the Mayan system of notation, the second position from the bottom corresponded to our (a) units, (b) tens, (c) hundreds, (d) thousands, place.

True–False

1. The ease with which multiplication can be carried out in the Roman numeral system was one of the principal reasons for the world's slowness in adopting the Hindu-Arabic notation.

2. Diderot was an outstanding mathematician of the seventeenth century.

3. In our number system the quotient of any number divided by zero is zero.

4. The Hindus are believed to have been the principal contributors to the development of the Hindu-Arabic system of numerals.

5. Neither the Greek nor the Roman number systems contained any symbol which was used as a place holder in the way that our notation employs the zero.

6. Neither multiplication nor division could be performed by the Romans.

7. Early man learned to use numbers abstractly long before he was able to apply them to specific objects.

8. The extensive use of multiples of five as bases for early number systems is believed to have been due largely to the fact that early man recognized only five planets.

9. CIV $-$ $\mu\alpha' =$ ⁝

10. $\frac{\bullet}{\text{ⅭⅠↃ}} + \nu\mu\beta' =$ CMII

Chapter 12

UNITS OF MEASUREMENT

The Role of Measurement

Lord Kelvin, a noted English scientist of the last century, once said, "When you can measure what you are talking about and express it in numbers, you know something about it. When you cannot measure it and express it in numbers, your knowledge is of a meager and unsatisfactory kind. It may be the beginning of knowledge but you have scarcely advanced in your thinking to the stage of a science." While the reader may or may not be inclined to accept this as an accurate test of knowledge, it is significant that in most fields of human endeavor the trend is constantly in the direction of more accurate measurement of more different quantities, and the expression of results in numerical form.

Certain of the scientific fields have advanced to such a point that they are called "exact sciences." In these the concepts involve quantities, most of which man is able to measure with a fair degree of accuracy. In other words, they are fields which lend themselves rather readily to the controlled, quantitative sort of investigation discussed in the first chapter. Other sciences involve things which are more difficult to measure, concepts and quantities that man is not yet capable of investigating experimentally with the same degree of quantitative control employed in the first group. These are not yet spoken of as exact sciences, but they are constantly moving in the direction of greater exactness, and greater accuracy of measurement.

The importance of measurement is not confined to the fields which we commonly call scientific. It affects almost every activity of our daily lives. One can illustrate this by tracing the events of a sort of composite average man's day. He may arise in the morning

from a bed whose length and breadth have been standardized so that mattress, springs, and bed will fit together, although manufactured in different localities. The innerspring mattress itself has in it a specific number of coils of wire of measured size, and with measured properties of elasticity, and fatigue. He may shave with a straight or safety razor whose properties are best for that purpose as a result of measured alloys, and measured conditions of temperature during manufacture. Or he may use an electric razor, the parts of which are so accurately made that they fit together within thousandths or tens of thousandths of inches. The clothing he wears, the food he eats, the news he reads, the house he lives in, the car he drives, in fact essentially all the activities of this typical man's typical day have had some contact with careful measurement.

Now measurement of any kind implies the existence of units in which the measurements can be made. It is certain of these that will be our primary concern in this chapter.

Fundamental Units of Measurement

Of the many kinds of quantities that we have occasion to measure in science, three are outstanding in an interesting and peculiar way. It happens that the units in which we measure almost all of the rest can be expressed in terms of one or more of the units of measurement of this favored trio. These three quantities are *length*, *mass*, and *time*, and their units are called " fundamental." The others are spoken of as " derived " and will be discussed and illustrated at the end of the chapter.

Units of Length

In the first place, it is necessary to understand clearly that the procedure which we call measuring length is always a matter of comparing one length with another. To say that a man is six feet tall is to say that he is six times as tall as a certain distance which we have agreed to call a foot. This is obvious, of course, but it involves two ideas which have been of immense importance in the development of units of all kinds.

Historical Background. — The first idea concerns the matter of accuracy of definition of the unit. Any length that we measure in

feet will be accurate only to the extent that we have been precise in defining a foot. The primitive man who was building a crude hut or dugout canoe had no need of extremely accurate units. Our modern world, on the other hand, in which, for example, one low priced automobile engine has over 30 points of contact in which tolerances of less than three ten-thousandths of an inch are permitted, requires very accurately defined units of measurement.

The second point has to do with the increase in standardization of units as time goes on. If the reader and his cousin in another state are to have the same idea of the height of a six-foot man, they must have the same definition of a foot. Here again is a matter that was unimportant to early man, since he had little to do with people at great distances from his own village. But it becomes exceedingly important in our present day civilization in which communication and transportation have made an actual fact of the ancient cliché, " It's a small world." It is essential that widely separated localities either use the same units, or define accurately the ones they do use in terms of the others.

It was natural that units of length should have their origin in the sizes of various parts of the human body. We may imagine one of our very early ancestors who wished to make a bow or spear exactly like the one possessed by his Paleolithic pal in the next village. Since his friend would hardly find it safe to lend him the weapon long enough to be used as a model, what would be more natural than that he should measure its dimensions in arm lengths, or hand breadths, and proceed to make a copy? As we have suggested, any lack of accuracy in such units was no particular drawback at this stage of civilization.

The records indicate that most of our units of length had some such anatomical origin. The Egyptian cubit was the distance from the elbow to the tip of the middle finger. The fathom was the length of the two arms outstretched. The yard was defined, at one time, as the distance from the point of the nose to the end of the thumb with the arm outstretched, and at another, as the distance around the waist. A span was the distance from the end of the thumb to the tip of the little finger, a palm the breadth of the four fingers, and the digit the breadth of the first finger. By checking on yourself you

can show that approximately, your cubit equals twice your span; your span is three times your palm; and your palm is four times your digit.

As soon as the necessity for greater accuracy and standardization began to be felt, we find the trend in these directions starting. Certain units came to be defined, either as averages or in terms of one specific individual or object. In 1324 Edward II of England decreed



FIG. 65. Early definition of the inch and foot. (Courtesy Ford Motor Co.)

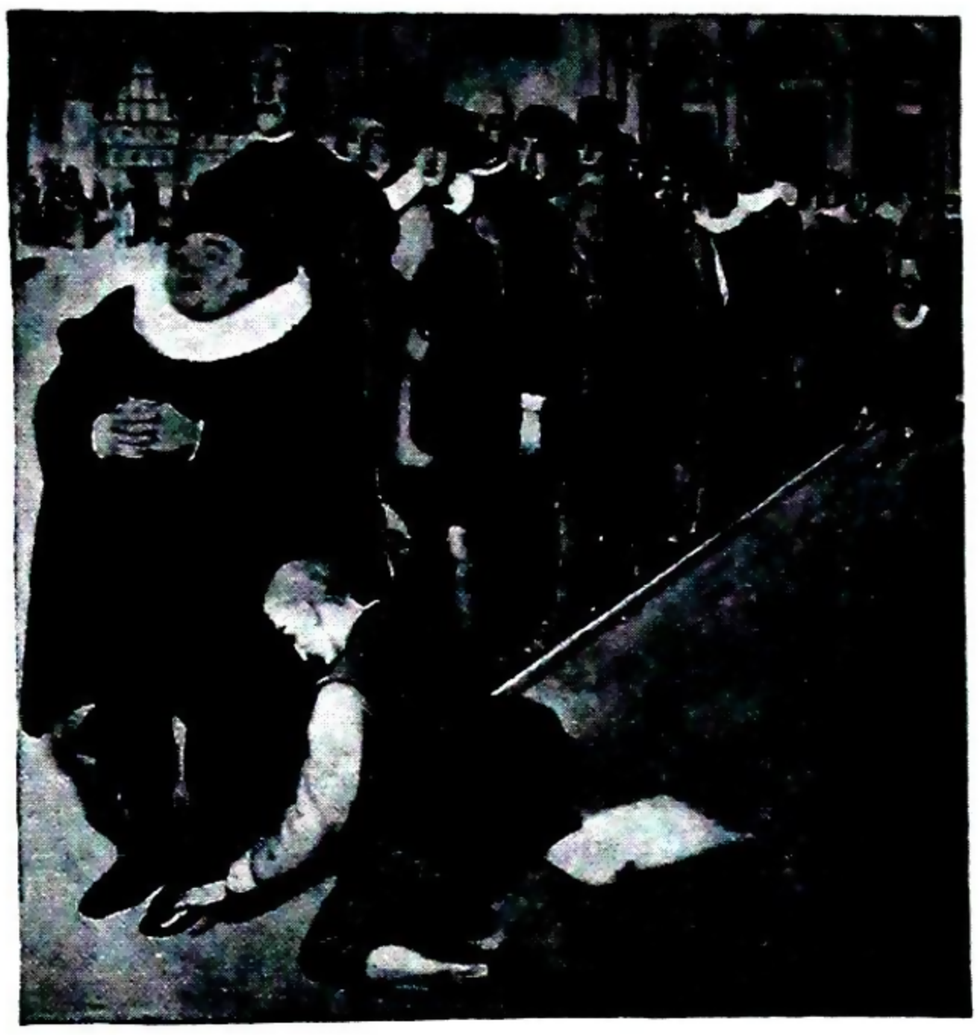


FIG. 66. Early definition of the rod. (Courtesy Ford Motor Co.)

that an inch should be the sum of the lengths of three barley grains taken from the middle of the ear (Fig. 65). By averaging over three grains a less variable inch would be established than if the definition had been simply three times the length of one grain of barley. A similar recognition of this averaging principle is found in a sixteenth century German rule for determining the length of a rod (Fig. 66). The instructions read, " Stand at the door of church on Sunday morning and bid sixteen men to stop, tall ones and small ones, as they leave the church; then make them put their left feet one behind the other, and the length thus obtained shall be a right and lawful rod to measure and survey land with, and the sixteenth part of it shall be a right and lawful foot." Here again was a definition which, although a little awkward, was considerably better than one which defined a rod as sixteen times the length of a man's foot. Henry I of England applied the other method of achieving accuracy of definition by decreeing

that a yard should be the distance from the tip of his own nose to the end of his own thumb when his arm was outstretched.

From these crude beginnings the measurement of length has advanced until today we have units that are well standardized and defined with great accuracy.

Commonly Used Units. — Of the two groups of length units in fairly common use today we are most familiar with the so-called English system, which includes the inch, foot, and yard. There are, however, some obvious advantages in a decimal system in which each unit is ten times the next smaller and one-tenth of the next larger. The metric system is such an arrangement with the meter being its fundamental unit of length. By attaching to the word "meter" a series

TABLE 11. METRIC PREFIXES AND THEIR APPLICATION TO UNITS OF LENGTH AND MASS

Prefix		*Length*	*Mass*
kilo	(1000)	kilometer	kilogram
hecta	(100)	hectameter	hectagram
deka	(10)	dekameter	dekagram
		meter	gram
deci	(.1)	decimeter	decigram
centi	(.01)	centimeter	centigram
milli	(.001)	millimeter	milligram

of prefixes having the meanings shown in parenthesis in the accompanying table, the various multiples and fractions of this standard distance are defined. The same prefixes may be used in the measurement of weight, area, volume, or any other quantity, once the fundamental unit has been defined. In the case of the meter it was originally intended that this basic unit should equal one ten-millionth of the distance from the equator to the North Pole. While such a basis may have had some theoretical advantages, it was a little inconvenient for practical comparison purposes. At the present time the fundamental definition of the meter is simply the distance between two scratches on a certain platinum-iridium bar kept at a particular temperature in the French bureau of weights and measures.

Legal Length Standard. — Although in the United States we use English units for most of our length measurements, it is a fact that for a number of years the meter has been our fundamental legal

standard. About 1893 Congress passed a law making the meter the legal standard for the United States and defining the yard as 3600/3937 of this meter. Since all other English units can be expressed as fractions or multiples of the yard, as soon as it was accurately defined, all were so defined. From the definition of the yard the reader will notice that it is just a little shorter than a meter, in fact just 3.37 inches shorter. The two units are compared in Fig. 67.

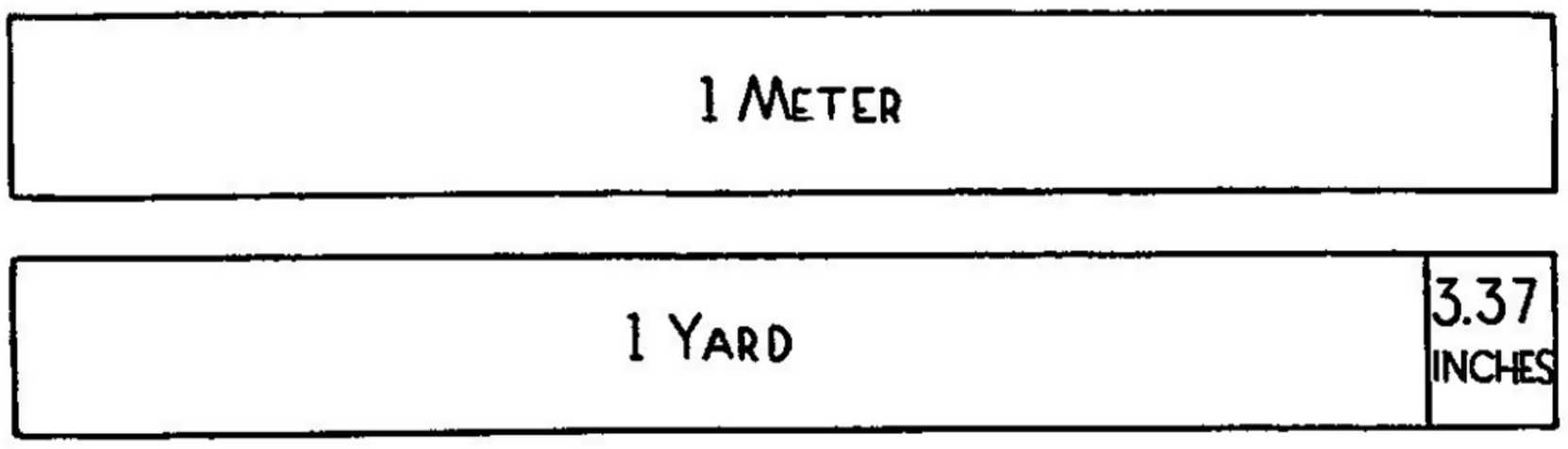


FIG. 67. Diagram showing comparative lengths of the meter and the yard.

A word may be said concerning the accuracy of length measurement that is possible today. We have already mentioned one example of the permissible tolerances in automobile engines. The acme of precision is probably exemplified by the C. E. Johansson Standard Gage Blocks. These gage blocks, manufactured by the Johansson Division of the Ford Motor Company, are rectangular pieces of tool steel used as length standards in the automotive and other industries. Several of the blocks are shown in use in Fig. 68. They are finished to guaranteed accuracies of one, two, four, or eight millionths of an inch, depending on requirements. One millionth of an inch is some 3,000 times finer than the thickness of a human hair or a sheet of newspaper. This illustrates the accuracy that is achieved in our present-day world and which makes possible the mass production of the precision machinery that we use every day.

Units of Mass

Historical Background. — Our modern units of mass have historical backgrounds similar to those of the length units. At first, weights of objects were simply compared with other objects which happened to be at hand. Stones with weights marked on them have been found in deposits dating back to 3400 B.C. Here again, as the need for greater accuracy and standardization arose, it was supplied

until today the precision that is possible in the measurement of weight compares favorably with that just discussed for length.

The English system of weight units includes the familiar ounce, pound, and ton. The ounce can be traced back to Rome where the weight of a cubic foot of water was taken as the standard unit. This

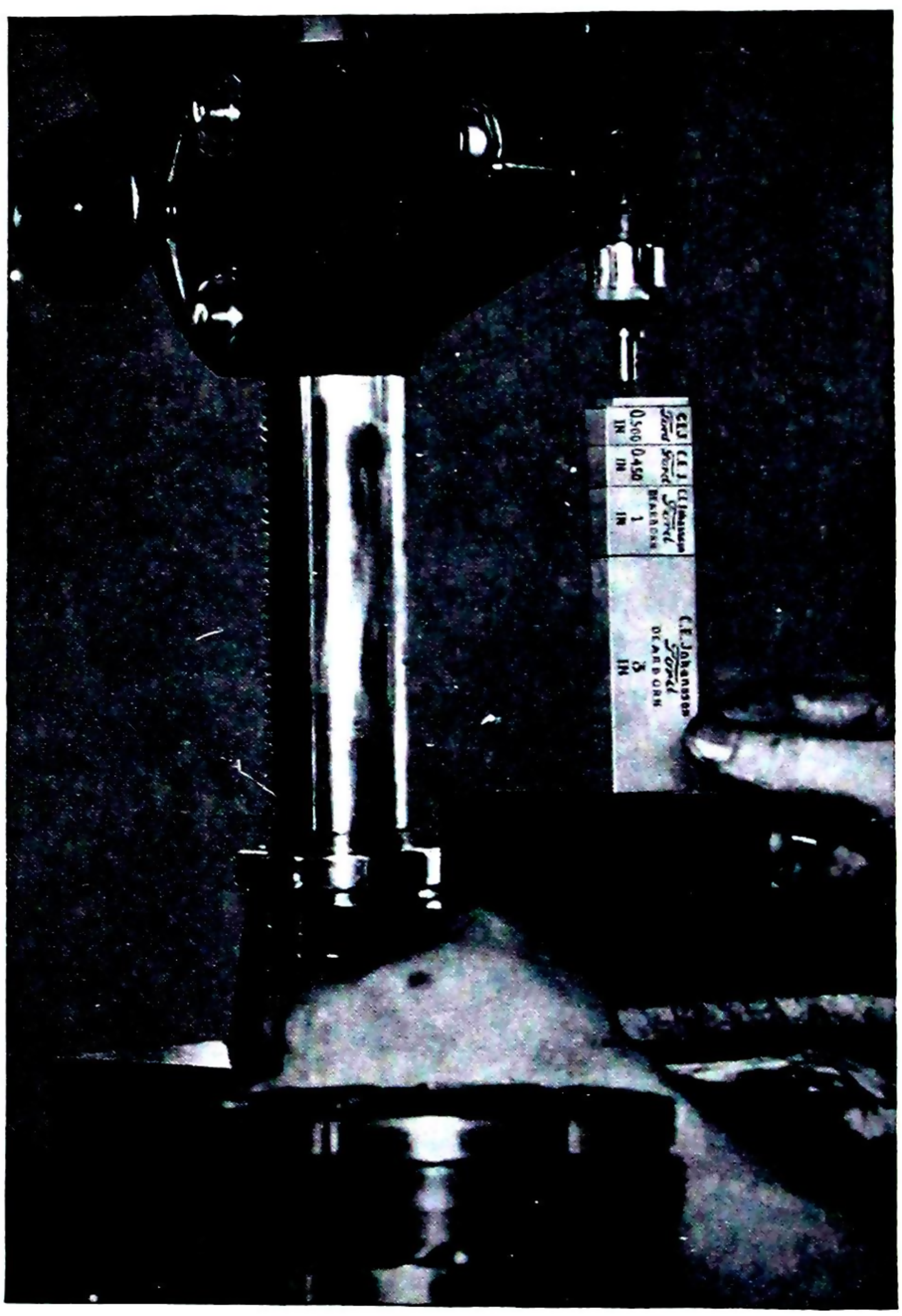


FIG. 68. Photo showing the Johansson Gage Blocks in use. (Courtesy Ford Motor Co.)

was divided into a thousand smaller units called ounces. The weight of a pint of wine was taken as the standard pound in early English units.

Commonly Used Units. — As has been suggested, the metric system of weights makes use of the same set of prefixes as described for length units, attaching them in this case to the gram. The gram

was intended to be the fundamental unit and was defined as the weight of a cubic centimeter of water at its temperature of greatest density. Now, however, the legal metric standard of weight is a platinum-iridium cylinder weighing one kilogram and located with the standard meter at the French bureau of weights and measures. The United States is on the metric system of weights also since Congress has defined a pound as 1/2.204622 kilograms. In other words, a kilogram equals approximately 2 1/5 pounds. Fig. 69 illustrates the relationship between the pound and the kilogram.

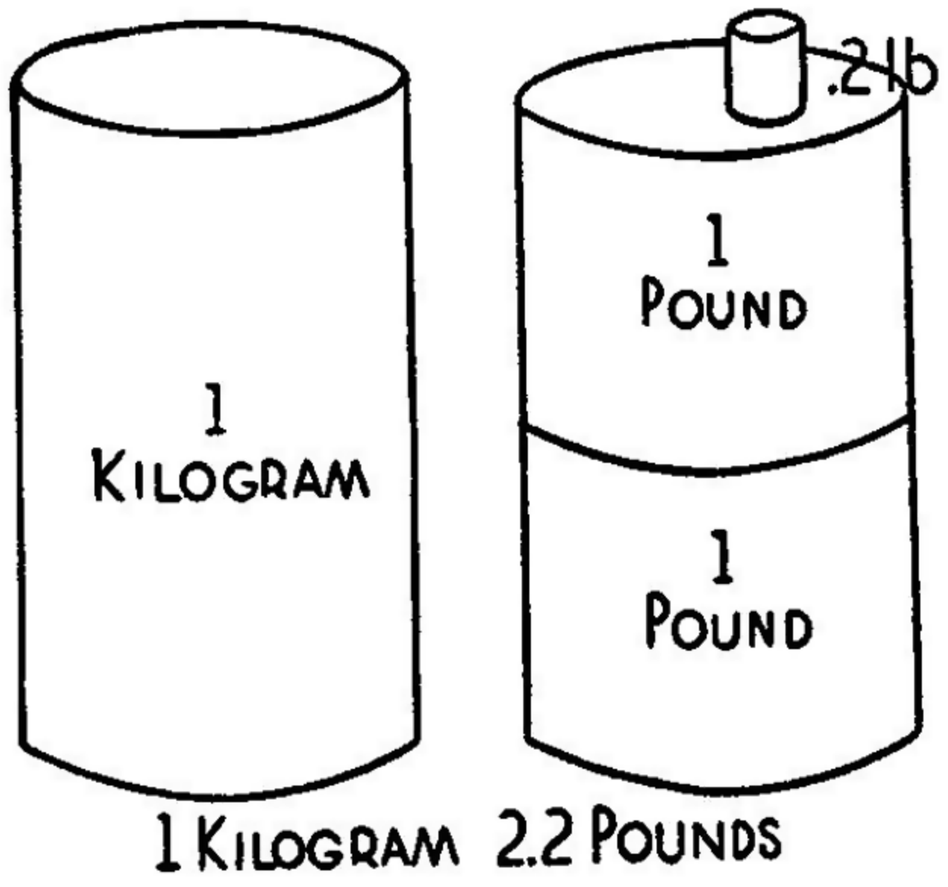


FIG. 69. Diagram illustrating comparative sizes of the kilogram and the pound.

Units of Time

The units in which we measure time are of two kinds, those determined by the relative motions of the sun, moon, and earth, which we may call natural units, and those which man himself has invented. The former include the day, month, and year, and the latter the week, hour, minute, and second.

Natural Units. — The first of the natural units to be recognized and used by man was undoubtedly the day, determined by the rotation of the earth on its axis.

The lunar month, which we no longer find very important, impressed itself on our primitive ancestors as the time required for the moon to go through its cycle of phases. Its length is approximately 29½ days.

The third of the natural units, the year, is established by the revolution of the earth about the sun and contains between 365 and 366 days. It was probably first marked off in terms of seasons or events such as summers, or winters, or snows or harvests. Somewhat later it was more accurately determined from observations made on the stars. It is a fact that any given star appears to rise a little earlier each night throughout the year. If we should fix our attention on the first star to rise after sunset tonight, we would find its time of

rising a little earlier each night until just one year from tonight when again, it would be the first to come up after sunset. By making such observations on Sirius, the priests of Egypt established a year of 365 days as early as 4200 B.C.

Earlier in this book we have mentioned one other natural unit of time, the Saros, which marks the interval between successive eclipses of the sun. It has no practical value to us since it is some 6585⅓ days, or 18 and a fraction years, in length.

Man-Made Units. — It should be said of all the man-made units of time that their lengths are quite arbitrary, since nature in no way marks them off. The idea of dividing days into groups smaller than a month, with a day of rest in each division, has been common among many peoples for many centuries, the number of days in a group varying from four to ten. Just where or when the seven-day week originated is not known. It came to us from the Romans but it was in existence before their civilization. The Hebrews have used it since early times but probably did not invent it.

Regardless of what people first used the seven-day week, it is almost certain that it originated because of worship of the seven heavenly bodies with which the ancients were familiar. The names which the days still have indicate the existence at some time of the belief that each of these celestial spheres ruled over the earth for one day out of every seven. Sunday, Monday, and Saturday suggest Sun's day, Moon's day, and Saturn's day. In Tuesday, Wednesday, Thursday, and Friday we have the days of the Saxon deities Tiw, Woden, Thor, and Frigg, corresponding to Mars, Mercury, Jupiter, and Venus respectively.

The use of the hour dates back at least to the Babylonians, since they were responsible for its division into minutes and seconds. But it was the Romans who were responsible for the spread of its use throughout the civilized world. The Roman hour differed from the one that we use today in that the daylight was divided into twelve equal parts, each called an hour, as was also the darkness. There were twenty-four hours in a day but they were varying hours which changed length with the seasons. This system seems to have worked fairly satisfactorily until the invention of mechanical clocks. It was too much to expect a clock to speed up and slow down to fit the varv-

ing amounts of daylight and darkness during the year, although there are records of a Japanese clock which partially solved the problem by having two faces, one for winter and the other for summer. A better solution proved to be the present practice of dividing the entire day into twenty-four equal hours. A day of this length is in almost universal use now, although others have been tried. The Chinese at one time had a ten-hour day. Following the French Revolution both the number of hours in a day and days in a week were changed, in line with the usual procedure after such political disturbances of assuming that " whatever was, was wrong."

The division of the hour into minutes and the minute into seconds was made by the Babylonians, to whom the number 60 had great significance.

Of these various divisions of time, the second is chosen as the scientific standard. A little multiplication will prove that there are 86,400 seconds in a day. Now if we think of the day as the time between successive passages of the sun across the meridian, we find that it varies a little during the year. Consequently, the day's length is averaged over a period of one year and the second is defined as " one 86,400th of the average time between successive passages of the sun across the meridian."

The Calendar

Man used the natural time units of day, lunar or moon month, and solar year for a good many generations before he made any attempt to put them together into a calendar. When one does try to devise such a system certain very troublesome situations arise. These come about largely because of the following relationships:

A lunar month equals 29 days, 12 hours, 43+ minutes;

A solar year equals 12.36 lunar months;

A solar year equals 365 days, 5 hours, 48+ minutes.

In these relationships a lunar month may be defined as the time between successive new moons, a solar year as the period of the earth's revolution about the sun, and a day as the period of the earth's rotation on its axis.

It will be noted that not one of the three units is an exact whole multiple of any other. The problem which this fact introduces when one attempts to devise a calendar may be illustrated as follows:

Suppose we should decide to set up a calendar using an exact solar year as our basic unit. If we start the first year at exactly midnight of a certain day, the second year will start 5 hours and 49 minutes after midnight of December 31. Also, if we wait to start our first year until we can have a new moon occur just at midnight, then at the end of the year we will have .36 of a month left over before the first new moon of the second year. Whichever one of the three units we might choose as basic, the others would cause trouble if we wanted day, month, and year all to begin at the same time. Before looking at our present calendar, let us see how some ancient peoples met these problems.

Early Calendars

The Babylonians were among the first nations to devise a calendar. Their early calendar was set up on the basis of a year of 360 days, divided into 12 months of 30 days each. It was intended that the months should be lunar months, each starting with a new moon. This was because the day of a new moon had important significance in various of their religious rites. Also, they believed that the day of a full moon was very evil and that to allow the light of such a moon to fall on one would affect the mind and produce the condition that came to be called lunacy. The practice of starting each month with a new moon apparently made it easier for people to figure just when these harmful full moon times would occur.

But a 30-day month is over 11 hours longer than the time between successive new moons. So occasionally it would happen that the new moon would arrive on the last day of a month instead of on the first day of the next one. When this was about to happen the priests announced that the thirtieth day would be omitted from that month, and in this way 29-day months made their way into the calendar.

Serious difficulties also came from the use of a 360-day year. Suppose we were to begin next January to put such a calendar into operation. The first New Year's Eve would come at the usual season. But since the earth would take no account of our streamlined year but would continue to take approximately 365¼ days to travel around its orbit, the second New Year's Eve would occur 5¼ days earlier in the winter. The third one would be 10½ days too early and so

on until after some 34 years we would be celebrating that particular holiday in the middle of the summer. Exactly this sort of thing began to occur with the Babylonians and, since their months and feast days were mostly named for certain seasonal events, the problem was a serious one. A citizen of Babylon, after a certain number of years, would find the Month of Plowing appearing on the calendar about harvest season, or the Feast of New Mown Hay coming in the middle of winter. Again the task of adjustment was delegated to the priests. Whenever they recognized that the situation was becoming serious, they announced that the current month would be repeated. It was as though at the end of some certain June, we would all agree that we liked June so well we would just hold it over for another 30 days. By this somewhat haphazard introduction of occasional years of 13 months, dates and seasons were kept approximately in their correct relationship.

The Greek calendar was similar to the Babylonian except that the extra months were inserted according to a definite procedure. The Greek astronomer Meton devised a rule that had appeared during the latter part of the Babylonian civilization. This system divided regular years into groups of nineteen called " great " years. By specifying seven of these as 13-month years, and the others as the 12-month years, proper relationships were maintained. The Jewish calendar in use today is based on a similar plan.

In the early Roman republic, the insertion of the extra months was again placed in the hands of the priests. Romans of that day were generally quite politically conscious and the priesthood was no exception. They recognized the advantages of having friendly officials in power and began to see to it that the men they liked held office for 13-month years while the others were allowed only the 12-month years for their tenure. After a few years of favoring politics over astronomy in the determination of extra-month years, the calendar was in hopeless confusion.

The Julian Calendar

The condition of the Roman calendar was so bad by the time Julius Caesar came to power that he decided to abandon it entirely and devise a new one. With the aid of Egyptian astronomers, a cal-

endar based on a year of 365 1/4 days was worked out. In order to avoid starting years at a time of day other than midnight, the practice of giving every fourth year an extra day and calling it a leap year was adopted. In the Julian calendar every year evenly divisible by four was such a year. The practice of beginning each month with a new moon was abandoned. The year was simply divided into 12 months without regard to the moon's phases. The number of days in each month as we now have them goes back to the Julian calendar where certain religious and political considerations entered into their determination. We also owe the names of the months to the Romans. January, March, April, May and June were named in honor of certain gods or groups of gods. February came from a religious festival. July and August obviously commemorate the Caesars, Julius and Augustus. September, October, November and December were the seventh, eighth, ninth and tenth months respectively at a time when the Romans began their calendar year with March instead of January.

The Gregorian Calendar

It will be noted that a year of 365 1/4 days is about 11 minutes toc long. Such an error was negligible over a period of a few years but became of some importance after several centuries had elapsed. By the sixteenth century it amounted to some 10 days, so that feast days which were established in the fourth century were just that much out of step with the seasons in 1500. This led to a revision of the Julian calendar by Pope Gregory XIII. Since the Julian calendar was a little too long, the error could be corrected by making ordinary years of some of the leap years. With the aid of Italian astronomers Pope Gregory worked out a new leap year rule which is the one that we still follow. It may be stated as follows:

Century years shall be leap years only if they are evenly divisible by 400; all others, if evenly divisible by 4.

That meant that, of the years 1600, 1700, 1800, 1900, only 1600 would be a leap year by the Gregorian calendar, although all would be by the Julian system.

In Roman Catholic countries the Gregorian calendar was put into operation in the fall of 1582. In order to get the calendar dates

back to their proper seasonal relationships, the 10 days from October 5 to October 14 inclusive were omitted. The English speaking countries continued to use the Julian calendar until 1752. By that time the discrepancy amounted to 11 days, so the English government ruled that the day after September 2, 1752, should be September 14. Although a great number of mass meetings were held at which people protested that their rulers were robbing them of 11 days, the plan went into effect on schedule.

The Gregorian calendar will maintain proper seasonal relationships until the year 4000 A.D. If one day is dropped out every four thousand years from that date on, it is believed that no serious difficulties will develop until about the year 20,000 A.D. Beyond that we perhaps need not concern ourselves just now.

Proposed Calendar Changes

From an astronomical standpoint, as we have seen, the present calendar is practically perfect. It now has the additional advantage of being in use in all the important countries of the world, Rumania and Turkey having adopted it since the close of the first World War.

At the same time it does have certain drawbacks and considerable objection has been made to it, particularly by business men. Months do not all have the same number of days in them. The same is true of half and quarter years. Weeks do not fit evenly into either months or years. These things make the comparison of business for different periods difficult, and introduce various problems in connection with salary and wage schedules.

Therefore the League of Nations appointed a committee to study the question of calendar reform. Of the almost 200 plans submitted, the following basic outline has received the most favorable comment. The proposed calendar year would consist of 13 months of 28 days each which would account for 364 days. The extra day, which would be a world holiday, would be a part of no week or month. Leap years would have two such holidays. Each month would have exactly four weeks in it; every month would start with the first day of the week; any given date would always fall on the same day of the week; and printed calendars would be the same every year. However, the adoption of such a plan by any large

number of nations seems unlikely in the very near future, and the troubles involved in changing over would not be worth while unless a considerable portion of the world did concur.

Derived Units of Measurement

Units of length, mass, and time have been described as being fundamental because of the fact that almost all other units can be expressed in terms of one or more of the three. The measurement of area provides an excellent example of this. We commonly express the area of a surface in such units as square inches, square feet, square miles. But the inch, the foot, and the mile are all units of length. Thus, to get an area unit, we have taken one of the fundamental units and squared it.

Similarly, to measure volume we use units such as cubic inches, cubic feet, and cubic miles, all of which are the cubes of fundamental units of measurement. Volume units like liters and gallons can be defined in terms of others that are cubes of lengths.

As an example of a measure that involves two of the fundamental units we may consider speed. One's automobile speedometer measures this quantity in terms of miles per hour. In other words, this unit of speed equals a fundamental unit, the mile, divided by another fundamental unit, the hour, or in general, length or distance divided by time.

This holds for most units of measurement. Consequently those of the latter type are called " derived " units to distinguish them from the three " fundamental " units from which they come.

Questions

Discussion

1. Define fundamental and derived units and give examples of both kinds.
2. Why is it more important today to define units of measurement accurately than it was one or two hundred years ago?
3. What difference would it make if the legal definition of a foot as used in California were to be an inch or so different from that employed by the people of New York?
4. Why was it better to define the rod as " the sum of the lengths of the left feet of sixteen men " than as " sixteen times the length of a man's left foot "?

5. Give both the original and the present-day legal definition of the following: (a) Yard; (b) Gram; (c) Meter.

6. List the prefixes employed in the metric systems of units and state the meaning of each.

7. Express your height in meters and your weight in kilograms.

8. Distinguish between natural and man-made units of time and give examples of each.

9. How would you define a lunar month? A calendar month?

10. Explain why the solar year, the lunar month, and the day do not all fit together into a satisfactory calendar.

11. Look up additional information concerning the early history of the week, the hour, the minute, and the second.

12. State the scientific definition of a second of time.

13. How did the Babylonians and the Romans meet the problems connected with making a calendar that would contain days, months, and years?

14. Distinguish between the Julian and Gregorian calendars.

15. What was the last year that was not a leap year but would have been on the Julian calendar, and when will the next one be?

16. Explain fully why it was necessary to omit all the days from September 3 to September 13 inclusive in 1752 when the Gregorian calendar was adopted in England?

17. Look up additional information on the disadvantages of our present calendar and on proposed new calendars and either discuss the question or write an essay on it.

Multiple Choice

1. Arrange the following possible definitions of a yard in order from poorest to best. (a) The distance from the governor's nose to the end of his thumb when his arm is outstretched. (b) Three times the length of just anyone's foot. (c) One-fifth the sum of the lengths of the left feet of fifteen men chosen at random. (d) An exact fraction of the length of a certain metallic bar. (e) Three times the average length of the left feet of any two men.

2. Of the following years, list all that would be leap years on the Julian calendar but not on the Gregorian. (a) 1066, (b) 1776, (c) 1400, (d) 1492, (e) 1700.

3. Butter at 40 cents per pound is — (a) much more expensive than, (b) a little more expensive than, (c) exactly the same price as, (d) a little cheaper than, (e) much cheaper than — butter at 80 cents per kilogram.

4. If next January, we should begin to operate on the basis of a calendar year of exactly 365 days; 100 years from now, the 4th of July would come (a) at the same season that it now does, (b) less than a month earlier in the summer than now, (c) in the middle of winter, (d) at about the season

that Labor Day now occurs, (e) about a month later in the summer than now.

5. Units of pressure are derived units; which one of the following combinations of fundamental units gives a commonly used unit of pressure? (a) Length × weight × time. (b) Length^2 / time^2. (c) Weight / length^2. (d) Weight / time^3. (e) Length × weight.

True–False

1. A million millimeters equals one kilometer.
2. A true solar year is a little less than 365 ¼ days in length.
3. The names of the months of the year can be traced back to the names of the heavenly bodies known to the ancients.
4. The quantity of gasoline that can be purchased for a dollar would be expressed in units which would be of the derived rather than the fundamental type.
5. A true lunar month is shorter than one-twelfth of the time it takes the earth to complete one revolution about the sun.
6. Of the four natural units of time, no one can be divided an exact whole number of times into any other.
7. Taken over a 400-year period, the average number of days in a year would be slightly less on the Gregorian calendar than on the Julian.
8. A six-foot man is more than two meters tall.
9. The gram was originally designed to be equal to the weight of one thousandth of a cubic foot of water.
10. In the metric system, the prefix " milli " means " one thousand."

Chapter 13

THE NATURE OF MATTER

THE first part of this book was concerned with man's acquisition of knowledge about the universe in general, and his own planet in particular. In these cases, we were considering phenomena over which man has no control. He can learn what the characteristics of the planets are and how they behave, what the changes in the earth's surface have been and probably will be, but nothing he is able to do can make them very different. He has discovered, though, that the methods of science can be applied in other fields to change the conditions under which he lives. And by applying these techniques he has greatly transformed his environment. We may not all agree on whether or not he has made the world a better place in which to live, but it can hardly be denied that he has made it different.

Consequently the rest of this book will concern the applications man has made of his scientific knowledge in the modification of his own environment. As discussed in the preface, these applications may be somewhat roughly divided into two groups. Man has learned much, on the one hand, about the composition and possible uses of actual substances, or materials, that is, the stuff of the world. On the other hand, he has enormously increased his knowledge of energy. He is learning more all the time about ways of using it, of changing it from one kind to another, and of getting it from one place to another. Although they overlap to some extent, these divisions can serve fairly satisfactorily as a basis for our discussion.

In the next few chapters we shall be concerned with what man has learned about materials. As we look about us we see hundreds of different substances. To discuss each individually, as something entirely separate and different from all the rest, would be a hopeless task. But we may speculate, as did our ancestors at a fairly early

date, as to the possibility of there being some sort of fundamental nature to all of them which would greatly simplify our studies.

We may illustrate the problem in the following fashion. Here, for example, are a theater, a cathedral, a business block, a dwelling house, a garage, and a hamburg stand, no one of which very closely resembles any other. Yet they may have the same basic nature in that they are all built of four kinds of fundamental building materials: brick, stone, steel and wood. In an analogous fashion, it was recognized at a fairly early time that it would be very desirable to discover that the multitude of objects and substances in the universe are all made of a comparatively few fundamental building blocks of nature. The reasons for believing that such fundamental blocks exist, the number of them, and how they behave will concern us in this chapter and the next one.

Early History of the Problem

Primitive Man's Knowledge

The relationship of primitive man to the world about him was a very practical one. Almost all of his time was required to supply his material wants of food, shelter, a certain amount of clothing, tools, and weapons. By methods of trial and error he learned to use certain materials which were easy to obtain, such as wood, skins, stone, and eventually, a few metals. It was many centuries before he became at all curious about questions of a more theoretical nature. As we have seen, his first such speculations concerned the stars and planets. Later he began to wonder about problems concerning the nature of matter, of heat, of light, and other phenomena.

Theories of the Greeks

Man soon recognized that there were just two possibilities with regard to the fundamental nature of matter. For one, it might be continuous. That is, one might be able to break a piece of iron in two, then divide one of the halves, again divide one of the quarters and so on indefinitely, never reaching a fundamental particle that could not be divided. Or, on the other hand, matter might be discontinuous or atomic in nature which would mean, in the above example, that one would eventually reach a fundamental particle

or atom that could not be further divided. In this case, any piece of iron or any quantity of matter at all would have to consist of an exact, whole number of atoms, just as a sand pile, regardless of size or shape, has in it an exact whole number of sand particles.

The Greeks seem to have been about the first people to theorize about this problem. Leucippus (about 460 B.C.) and Democritus (460–370 B.C.) were among the earliest proponents of an atomic theory. Many of their ideas, although based on little or no actual experiment, have proved quite sound in the light of our modern knowledge. Sir William Dampier states,[1] "Whatever may be its value in philosophy, in science the Democritean atomic theory is nearer to the views now held than any of the systems which preceded or replaced it, and its virtual suppression under the destructive criticism of Plato and Aristotle must, from the scientific standpoint, be counted a misfortune."

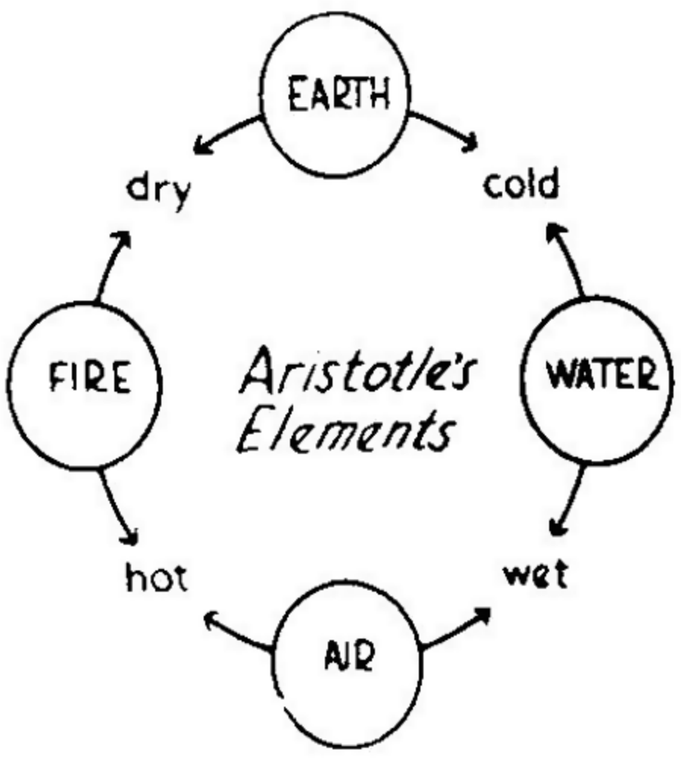


FIG. 70. Diagram illustrating the relationship taught by Aristotle to exist between the four elements and the four attributes.

Aristotle (384–322 B.C.) rejected the atomic theory because certain of its conclusions did not agree with some of his philosophical speculations on other aspects of nature. As a result of his great prestige, the idea was generally abandoned for the time being. He did, however, advocate a simplification of matter in that he believed all forms could be reduced to a few elemental kinds. By processes of reasoning which were supported by no experimental evidence whatsoever, he arrived at the conclusion that all matter in the universe was composed of four elements: fire, earth, air, and water. These he linked with four qualities or attributes of hotness, coldness, wetness and dryness, in the manner shown in Fig. 70. There was such a beautiful unity about this picture that it seemed as though it just couldn't be wrong. Therefore to everyone for the next 1800 years, air was hot and wet, fire was hot and dry, earth was dry and cold, and water was cold and wet. As a matter of fact, such a picture, erroneous as it was, did represent a step in the right direction, since it was an attempt to reduce all forms of material substance to a comparatively few fundamental kinds.

[1] Dampier, Sir William, *A History of Science*, p. 27. New York: Macmillan, 1935.

Revival of an Atomic Theory

Galileo adopted the general idea of an atomic nature of matter as fitting in well with his experiments on moving bodies. The form of the theory was revised by the physicist, Gassendi, around 1660, so that at that time matter was pictured as composed of small, hard, spherical particles called atoms. Almost nothing was specified as to the properties of the atoms themselves.

The atomic hypothesis was little more than speculation until the quantitative nature of various chemical reactions was established during the eighteenth century. A series of experiments by Bergman, Black, Cavendish, and Lavoisier during the period from 1750 to 1800 formed the basis of the atomic picture advanced by the English schoolmaster, John Dalton (1766–1844), in 1807. Since Dalton's basic postulates did not greatly differ from those of the present kinetic theory of matter and heat, we will proceed to a discussion of the experimental background of the latter.

Experimental Facts Leading up to Our Present Picture of Matter

The story of our present picture of matter may be thought of as falling naturally into two episodes. The complete history of a brick house would involve the construction of the house from bricks and the manufacture of bricks from simpler constituents. Similarly the first part of the story of the present accepted theory of matter deals with bulk matter being composed of atoms. The second episode, based on the revolutionary discoveries of the last few decades, deals with conclusions about the constitution of the atom itself. Since this second part involves certain electrical phenomena, our discussion of it will have to wait until a later chapter. Only the first episode will concern us at present.

The set of postulates which we believe, at the present time, gives us a fair description of the atomic nature of matter is known as the "Kinetic Theory." As we shall see, it involves both matter and heat. Our approach to it here, as far as possible, will be along the route prescribed by the scientific method. First we shall discuss a representative group of experimental facts about matter, that is, things we know are true. Then we shall set up the basic assumptions of

the kinetic theory and ask, " Does this theory account in a reasonable fashion for the experimental facts we have discussed? " It must be remembered that any scientific theory is satisfactory exactly to the extent that it does account for experimental phenomena and no further. To answer the above question completely, we would, of course, have to treat the numerical relationships between phenomena. Since such a quantitative discussion would lead us into mathematics beyond the scope of this book we will confine ourselves here for the most part to the qualitative aspects of the question and ask the reader to accept on faith that quantitatively also there is reasonable agreement. It should be mentioned that this procedure carries with it the danger of over-simplification. The reader should, therefore, be warned against thinking that all the difficulties have been solved and that our theories of matter are perfect. Fortunately there are still imperfections and problems to be solved in connection with most of them. Otherwise science would indeed be dull.

Kinds of Matter

By the time Dalton advanced his atomic theory of the nature of matter the existence of several different kinds of substances was recognized. Some knowledge of the principal characteristics of these types is essential to an understanding of other experimental facts about matter.

Elements. — Before stating a definition of an element that will be satisfactory here, let us see how these materials were defined by John Dalton. He had discovered that there were a number of substances which defied all his efforts to break them down into simpler materials. He reasoned, however, that if he were to attempt to subdivide one of them indefinitely, in the manner described earlier in the chapter for a piece of iron, he would eventually get a smallest possible particle of the material, one which could not be further decomposed. This ultimate particle of an element he called an atom. Consequently he would have defined an element as a substance that could not be decomposed into simpler ones. It happens, however, that within comparatively recent times, atoms of almost every known element have been broken up. This has been done in the so-called " atom-smashing " experiments that have received a considerable

amount of publicity. Some discussion of the methods employed in these experiments and of their relation to atomic energy occurs in Chapter 26 in connection with what we have called the second episode in the story of matter. Since one of the results of these investigations is that atoms are decomposed into simpler things, it is clear that Dalton's definition of an element must be modified. For the purposes of this discussion we can state it as, " An element is a substance which cannot be broken down into simpler materials *without* disrupting the atom."

John Dalton knew of about 20 different elements in 1807. Today we know of some 92, of which, at ordinary pressures and temperatures, two are liquids, some ten are gases, and the rest are solids. It has already been stated that 61 have been found in the sun, about 30 in meteorites, and many on other stars and planets. It is important to note that there is no evidence of the existence anywhere in the universe of elements other than these. Therefore it seems to be true that every one of the hundreds of thousands of different materials of which we have knowledge is a combination of one or more of not over 92 building blocks.

Now let us turn to Table 12. Here the elements are listed in alphabetical order in the first column. In the second column the chemical symbol for each element is given. It should be emphasized that these symbols are more than just abbreviations for the elements. Each one has a definite quantitative significance, representing quite definitely one atom of the element concerned.

Atoms of the elements have been found to have definite weights. We saw back in Chapter 12 that the process of weighing anything is simply one of comparing its weight with that of a recognized standard unit. In the case of butter or meat that unit is usually the pound. But, because atoms weigh so little, it is desirable to define a different standard unit. Originally this standard was established by giving hydrogen, the lightest element, an atomic weight of one and expressing the others in relation to it. Today oxygen is considered to be a more satisfactory standard and the oxygen atom is arbitrarily given a weight of 16. The actual unit of atomic weights is 1/16 the weight of an oxygen atom.

TABLE 12. ALPHABETICAL LIST OF THE ELEMENTS

Element	*Symbol*	*Atomic Weight*	*Atomic Number*
Aluminum	Al	26.97	13
Antimony	Sb	121.76	51
Argon	A	39.94	18
Arsenic	As	74.91	33
Barium	Ba	137.36	56
Beryllium	Be	9.02	4
Bismuth	Bi	209.00	83
Boron	B	10.82	5
Bromine	Br	79.916	35
Cadmium	Cd	112.41	48
Calcium	Ca	40.08	20
Carbon	C	12.00	6
Cerium	Ce	140.13	58
Cesium	Cs	132.91	55
Chlorine	Cl	35.457	17
Chromium	Cr	52.01	24
Cobalt	Co	58.94	27
Columbium	Cb	92.91	41
Copper	Cu	63.57	29
Dysprosium	Dy	162.46	66
Erbium	Er	167.64	68
Europium	Eu	152.0	63
Fluorine	F	19.00	9
Gadolinium	Gd	157.3	64
Gallium	Ga	69.72	31
Germanium	Ge	72.60	32
Gold	Au	197.2	79
Hafnium	Hf	178.6	72
Helium	He	4.002	2
Holmium	Ho	163.5	67
Hydrogen	H	1.0078	1
Indium	In	114.76	49
Iodine	I	126.92	53
Iridium	Ir	193.1	77
Iron	Fe	55.84	26
Krypton	Kr	83.7	36
Lanthanum	La	138.92	57
Lead	Pb	207.22	82
Lithium	Li	6.940	3
Lutecium	Lu	175.0	71
Magnesium	Mg	24.32	12
Manganese	Mn	54.93	25
Mercury	Hg	200.61	80

Element	*Symbol*	*Atomic Weight*	*Atomic Number*
Molybdenum	Mo	96.0	42
Neodymium	Nd	144.27	60
Neon	Ne	20.183	10
Neptunium	Np	239	93
Nickel	Ni	58.69	28
Nitrogen	N	14.008	7
Osmium	Os	191.5	76
Oxygen	O	16.0000	8
Palladium	Pd	106.7	46
Phosphorus	P	31.02	15
Platinum	Pt	195.23	78
Plutonium	Pu	239	94
Potassium	K	39.096	19
Praseodymium	Pr	140.92	59
Protactinium	Pa	231	91
Radium	Ra	226.05	88
Radon	Rn	222	86
Rhenium	Re	186.31	75
Rhodium	Rh	102.91	45
Rubidium	Rb	85.44	37
Ruthenium	Ru	101.7	44
Samarium	Sm	150.43	62
Scandium	Sc	45.10	21
Selenium	Se	78.96	34
Silicon	Si	28.06	14
Silver	Ag	107.880	47
Sodium	Na	22.997	11
Strontium	Sr	87.63	38
Sulfur	S	32.06	16
Tantalum	Ta	180.88	73
Tellurium	Te	127.61	52
Terbium	Tb	159.2	65
Thallium	Tl	204.39	81
Thorium	Th	232.12	90
Thulium	Tm	169.4	69
Tin	Sn	118.70	50
Titanium	Ti	47.90	22
Tungsten	W	184.0	74
Uranium	U	238.14	92
Vanadium	V	50.95	23
Xenon	Xe	131.3	54
Ytterbium	Yb	173.04	70
Yttrium	Y	88.92	39
Zinc	Zn	65.38	30
Zirconium	Zr	91.22	40

A much more important way of listing the elements is shown in Table 13. In 1863 an English chemist named Newlands happened to arrange the elements known at that time in order of increasing atomic weights. In studying his list he noticed that for the most part each element resembled closely, in physical and chemical properties, the eighth element following it. He announced this discovery, which he called his " Law of Octaves," but the exceptions were so numerous that his work was never taken very seriously. Then some six years later the Russian chemist Mendeleef stated his Periodic Law to the effect that, " the physical and chemical properties of the elements are periodic functions of their atomic weights." It is his arrangement of the elements in tabular form which constitutes the Periodic Table, shown in Table 13, a classification that has been called one of the greatest generalizations of all science. A few facts concerning the table should be mentioned.

It will be noted that in general atomic weights increase reading from left to right across the table. There are several exceptions to this, but the table is consistent in that all the elements are arranged from left to right along the rows, taking the rows from top to bottom, in order of increasing " atomic number." An exact definition of atomic number will have to await our discussion of the inside of the atom in Chapter 26. It should be said here, though, that using number instead of weight as the basis of arrangement serves to eliminate most of the inconsistencies that were present in Newland's table and in Mendeleef's earliest ones. When the elements are placed in this fashion, those in the same vertical column in general have similar properties. One important function of the Periodic Table has been as an aid in the discovery of new elements. When Mendeleef first prepared it, the table contained a number of blank places which were presumed to represent elements not yet discovered. The search for such an absentee was aided greatly by the fact that some idea of its properties could be obtained by studying the other members of the same vertical column. Similarly, by noting the atomic weights of those just preceding and following it in the horizontal row, an approximate determination of its weight could be made. The search for the missing members of the element family was simplified and many of the gaps were filled rather quickly.

The italic number at the right of the symbol is the Atomic Number of the element, and the number below is the Atomic Weight to one decimal place"

Periods		Group 0	Group I	Group II	Group III	Group IV	Group V	Group VI	Group VII	Group VIII
Formula of highest oxide...		—	R_2O	RO	R_2O_3	RO_2	R_2O_5	R_2O_6	R_2O_7	RO_4
Formula of hydride.....		—	RH	RH_2	RH_3	RH_4	RH_3	RH_2	RH	—
			H *1* 1.008							
First short period.....		He *2* 4.003	Li *3* 6.940	Be *4* 9.02	B *5* 10.82	C *6* 12.010	N *7* 14.008	O *8* 16.000	F *9* 19.00	
Second short period.....		Ne *10* 20.183	Na *11* 22.997	Mg *12* 24.32	Al *13* 26.97	Si *14* 28.06	P *15* 30.98	S *16* 32.06	Cl *17* 35.457	
First long period	Even Series..	A *18* 39.944	K *19* 39.096	Ca *20* 40.08	Sc *21* 45.10	Ti *22* 47.90	V *23* 50.95	Cr *24* 52.01	Mn *25* 54.93	Fe *26* Co *27* Ni *28* 55.85 58.94 58.69
	Odd Series...		Cu *29* 63.57	Zn *30* 65.38	Ga *31* 69.72	Ge *32* 72.60	As *33* 74.91	Se *34* 78.96	Br *35* 79.916	
Second long period	Even Series..	Kr *36* 83.7	Rb *37* 85.43	Sr *38* 87.63	Y *39* 88.92	Zr *40* 91.22	Cb *41* 92.91	Mo *42* 95.95	Ma *43*	Ru *44* Rh *45* Pd *46* 101.7 102.91 106.7
	Odd Series...		Ag *47* 107.880	Cd *48* 112.41	In *49* 114.76	Sn *50* 118.70	Sb *51* 121.76	Te *52* 127.61	I *53* 126.92	
Third long period	Even Series..	Xe *54* 131.3	Cs *55* 132.91	Ba *56* 137.36	La *57* 138.92	Ce *58* 140.13				
	Odd Series...			The Rare Earth Elements Pr *59* to Yb *71*						
Fourth long period	Even Series..					Hf *72* 178.6	Ta *73* 180.88	W *74* 183.92	Re *75* 186.31	Os *76* Ir *77* Pt *78* 190.2 193.1 195.23
	Odd Series...		Au *79* 197.2	Hg *80* 200.61	Tl *81* 204.39	Pb *82* 207.21	Bi *83* 209.00	Po *84*	—*85*	
Fifth long period		Rn An Tn *86* 222	—*87*	Ra *88* 226.05	Ac *89*	Th *90* 232.12	Pa *91*	U *92* 238.07		

Another value that the table had was the fact that its implications seemed to suggest the advisability of increased investigation of the inside of the atom. Similarities in properties of atoms indicated the possibility of similarities in internal structure rather than the existence of 92 entirely different and unlike building blocks of nature.

To summarize, then, the simplest of the kinds of matter we are discussing here is the element and the smallest division of an element is an atom.

Compounds. — There are several kinds of matter that *can* be more or less easily decomposed into simpler substances. The first of these to concern us is the group known as compounds. As an example we will consider the one compound for which practically everyone in the world knows the chemical formula, that is, water or H_2O. This formula means that water is composed of hydrogen and oxygen in the ratio of two atoms of the former to one of the latter. Now if we were to take a quantity of water and divide it again and again, as described previously for iron, we would eventually reach a point where there remained just two atoms of hydrogen and one of oxygen. This least possible amount of water is called a *molecule*. Any further division would result in our having, not water, but separate atoms of the simpler substances, hydrogen and oxygen. Similarly, table salt has the formula NaCl, which means that one atom of sodium (Na) and one atom of chlorine (Cl) are combined in one molecule of salt. It will be obvious that a chemical formula has just as specific a meaning as does a symbol, standing for one molecule of a compound in the above examples.

Perhaps the outstanding property of compounds is the fact that they possess definite composition. In other words, every molecule of a given compound has in it the same number of the same kinds of atoms as every other molecule. Water is always H_2O and table salt is always NaCl. We do not sometimes find water which is H_3O or HO_2, or salt represented by Na_2Cl, or Na_6Cl_7, or some other combination. This is important as a distinguishing feature of compounds.

Now if nature has only 92 different elements or building blocks with which to work, it is obvious that many changes must take place during the formation and decomposition of the great number of substances we see about us. Such transformations as concern the

changing of substances into other substances are called chemical changes or reactions. It is desirable that we acquire at least a passing acquaintance with the shorthand method by which such reactions are described.

But first, as a crude example, let us consider the change that occurs in a freight train when it is entirely uncoupled, or decomposed into its constituents. One might write a comparatively long paragraph describing the results of detaching from each other all the cars of our hypothetical train. On the other hand, one might take care of the essentials of the phenomenon by writing simply,

$$\text{Freight train} = \text{Engine} + 17\ \text{Cars} + \text{Caboose}$$

We have simplified our description by using that most terrifying of bugaboos, an equation. But, like most bugaboos, closer examination shows it to be quite harmless. This one is simply a more efficient way of describing something that happened, in this case the breaking up of a freight train.

Already, the reader will have realized that we have been attempting, as it were, to " sneak up on him " with a chemical equation, for a chemical equation performs exactly the same function for a chemical reaction as the above equation did for the railway yard occurrence. Suppose, for example, we set out to describe the results of passing an electric current through water. We find *experimentally* that the reaction which occurs results in decomposing the water into its constituents, hydrogen and oxygen. Thus at first thought we might write,

$$H_2O = H + O,$$

This can be strongly objected to on two counts. In the first place, it isn't an equation. The very word implies equality and here we have two atoms of hydrogen on one side of the equality sign and only one on the other. The second objection is that it does not tell a complete, true story. There are some gases in which, under normal conditions, the atoms, like people at a dance, group themselves in pairs. These are said to be diatomic (two atom) gases. Hydrogen and oxygen belong to this group. Other gases, such as helium, contain atoms which are of a more independent nature and are said to be monatomic

(one atom). To represent accurately this experimental fact, the hydrogen and oxygen in our experiment must be written as H_2 and O_2. Making this correction in our alleged equation we have,

$$H_2O = H_2 + O_2.$$

But it still isn't an equation. The other objection can be eliminated by taking more water to begin with. Thus, when the proper coefficients are inserted, a process known as balancing, we get,

$$2H_2O \rightarrow 2H_2 + O_2.$$

In one short line this expression correctly describes the results of our experiment, both qualitatively and quantitatively. To say the same thing and avoid an equation we would have to write something like " when an electric current is passed through water, the water is decomposed into its constituents, hydrogen and oxygen, in a manner such that from every two molecules of water, each molecule being composed of two atoms of hydrogen and one of oxygen, there are obtained two molecules of hydrogen, each of which consists of two atoms of hydrogen, and one molecule of oxygen having in it two oxygen atoms."

As one more example, suppose we wish to describe the reaction that occurs when hydrochloric acid is poured on aluminum. Experimentally, we find that a gas, hydrogen, is given off, and that there remains in the test tube a substance which proves to be a compound of the elements aluminum and chlorine, called aluminum chloride. The equation which completely describes this chemical occurrence is

$$2Al + 6HCl \rightarrow 3H_2 + 2AlCl_3$$

which simply says that,

2 atoms of aluminum + 6 molecules of hydrochloric acid → 3 molecules of diatomic hydrogen + 2 molecules of aluminum chloride

Summarizing here then, a chemical equation has the primary function of describing the results of a given chemical reaction. To be satisfactory it must actually be an equation, i.e. be balanced, and must tell as complete a story as possible.

Other Kinds of Matter. — Three other combinations or types of matter should be mentioned. *Mixtures* are exemplified by the sort

of thing one would get by stirring dry sand and sugar together. As the name implies, merely a physical mixing up of the constituents occurs with no chemical reactions involved, either in making the mixture or in decomposing it.

Solutions, on the other hand, constitute a somewhat closer combination of the substances involved, as for example, sugar dissolved in water. The combination is not such that a compound is formed since the property of definite composition is not present. In the example mentioned, for instance, a wide variety of amounts of sugar might be dissolved in the same volume of water.

A group of substances known as *Colloids* will be discussed in Chapter 18.

States of Matter and Some Properties of Each

In the above paragraphs we have considered certain facts about the kinds of matter that exist. Now, let us remind the reader of our aim in this chapter and the next as expressed earlier. We are engaged now in examining some typical experimental facts about matter with the purpose of eventually arriving at a theory of its nature which will be in agreement with the observed data. Many of the familiar facts about substances have to do with the three states in which matter exists, namely as solids, liquids, and gases.

Solids. — The first thing to be noted about solids is the characteristic by which we commonly distinguish them from the other two states, that is, the fact that they are able to maintain their shape in three dimensions. They neither flow as do liquids, nor expand and fill all available space as do gases. This immediately suggests to us that there must be fairly large forces of some sort holding the molecules together. That these forces are very large may be seen from the fact that it requires some 25 tons to pull apart a bar of rivet iron of 1 square inch cross-section.

Solids also exhibit very great resistance to compression. In fact it is almost impossible to compress them at all. At first this might lead us to assume that the molecules must be packed tightly against each other. However, experiments show us that solids can be made to diffuse slightly into each other. For example, if a block of lead and a block of gold are bound tightly together, the contact surfaces

having been shaved clean, and left for some time, flecks of gold will be found to have penetrated the lead while bits of lead may be seen in the gold. Since this could scarcely happen if molecules are jammed together, we are forced to assume that even in solids, the molecules are separated by spaces. This same evidence would lead us also to suspect that the molecules are in motion. Already, you see, we are beginning to make assumptions about the nature of matter, assumptions that must be checked by other evidence.

Liquids. — Liquids neither maintain shape as do solids nor fill all available space as do gases. They adopt a sort of middle-of-the-road policy, flowing to fit the shape of the container but at the same time maintaining what may be called a free surface. This would seem to suggest that the forces of attraction between the molecules are appreciable but too small to overcome the force of gravity tending to make the liquid flow.

Liquids resist compression almost as much as do solids. If this were not true, we would have no such devices as hydraulic brakes, elevators, or barber chairs. You would not care to have hydraulic brakes on your automobile if the liquid itself was easily compressed when you pushed on the brake pedal.

Liquids are found, in general, to diffuse more readily than solids but much less easily than gases. If our assumption about space between molecules in the case of solids was correct, this would suggest that in liquids the separation is still greater.

The familiar phenomenon of evaporation must be considered in our picture of liquids. The two experimental facts of importance are first, that liquids do evaporate, and second, that the rate of evaporation increases with the temperature. It is common knowledge that, other things being equal, clothes dry faster on warm days than on cold ones, and that the water in the humidifier of the furnace evaporates more rapidly with a fire in the furnace than without one. These facts also must be made to appear reasonable in the light of any satisfactory theory which we may advance.

Gases. — Gases, of course, do not even maintain a free surface but expand and fill all the space available. A new tank of oxygen is full, and it is still full after half or after nine-tenths of the gas has been used. This would seem to indicate that the forces be-

tween molecules which were large in the case of solids, smaller but appreciable in liquids, have now become so small as to be negligible.

In the matter of diffusion we find that gases exhibit the phenomenon to a much greater extent than do either solids or liquids. A little perfume set free in one corner can soon be detected throughout the room. The vapor from onions or cabbage cooking in the kitchen seems to experience little difficulty in moving through and with the atmosphere to the front door. In line with our tentative conclusions from this phenomenon in solids and liquids, we would suspect that comparatively large separation exists between molecules of gases.

The question of compression in gases is tied up with a group of relationships called the gas laws. They are a set of three experimental laws relating pressure and volume, pressure and temperature, and volume and temperature, respectively, of an enclosed gas. For our purposes it will be sufficient if we consider the first two of these.

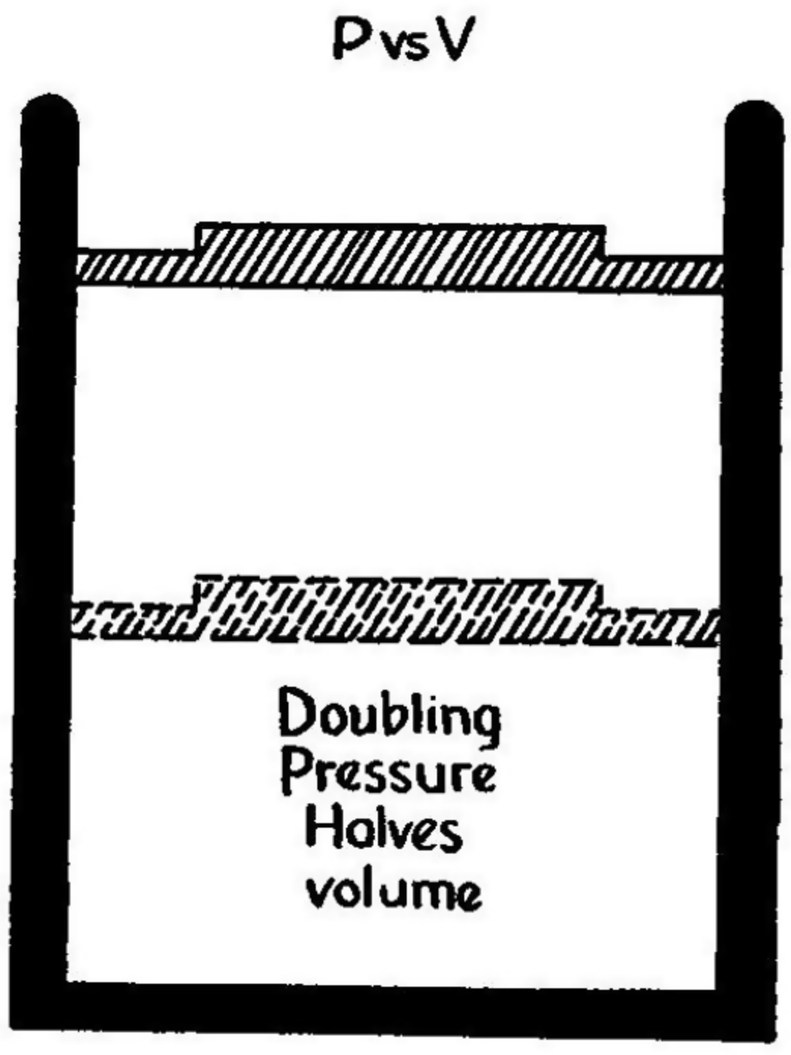


FIG. 71. Diagram illustrating the experimental facts expressed by Boyle's law.

Everyone is familiar with the fact that gases are easily compressed or confined to smaller volumes. Fig. 71 shows a cylinder in which there is an airtight plunger or piston. The air or other gas in the cylinder occupies a certain volume and exerts a certain pressure on the piston as well as on the sides and bottom of the cylinder, just as the air in an automobile tire may exert a pressure of 30 or 35 pounds per square inch above atmospheric pressure. Now, as to the question of the relationship between the pressure and volume, it is obvious that if we increase the pressure on the piston, it will move down the cylinder and the volume of gas will be diminished. In other words, as one of these factors gets larger, the other becomes smaller. Such a statement is qualitative. Careful experiment shows that numerically, or quantitatively, the relationship is what is known as an *inverse proportion*. This means that if the pressure on the piston is doubled,

the volume becomes exactly one-half as great; or if the pressure is tripled the volume is reduced to one-third its original value. These experimental facts are stated in what is known as *Boyle's law*, as follows:

> *If the temperature is kept constant, the volume of an enclosed mass of gas varies inversely with its pressure.*

The second gas law relates pressure and temperature of a similarly enclosed mass of gas when the volume remains the same. Consider the airtight cylinder shown in Fig. 72. It is filled with air or other gas at room temperature. This gas exerts a certain pressure on the inside of the cylinder, a pressure which is indicated on the attached gauge. Now it is a matter of common knowledge that if we were to place this cylinder on the stove and heat the gas in it, the pressure would become greater. In this fashion pressure is built up in a pressure cooker, or in automobile tires on a hot day. In other words we have here two factors, pressure and temperature, so related that as one gets larger the other also increases. Again, careful experiment would show the numerical relationship between the two to be what is known as a *direct proportion*. This means that the pressure will be doubled if the temperature is doubled, or tripled if the temperature becomes three times as large. It must be stated that the temperature referred to here is one based on neither the Fahrenheit nor Centigrade scales, but on what is known as the Absolute scale. This basis of temperature measurement will be defined and discussed in the next chapter. The relationship between pressure and temperature of a gas bears the name of the *Gay-Lussac law* and may be stated as follows:

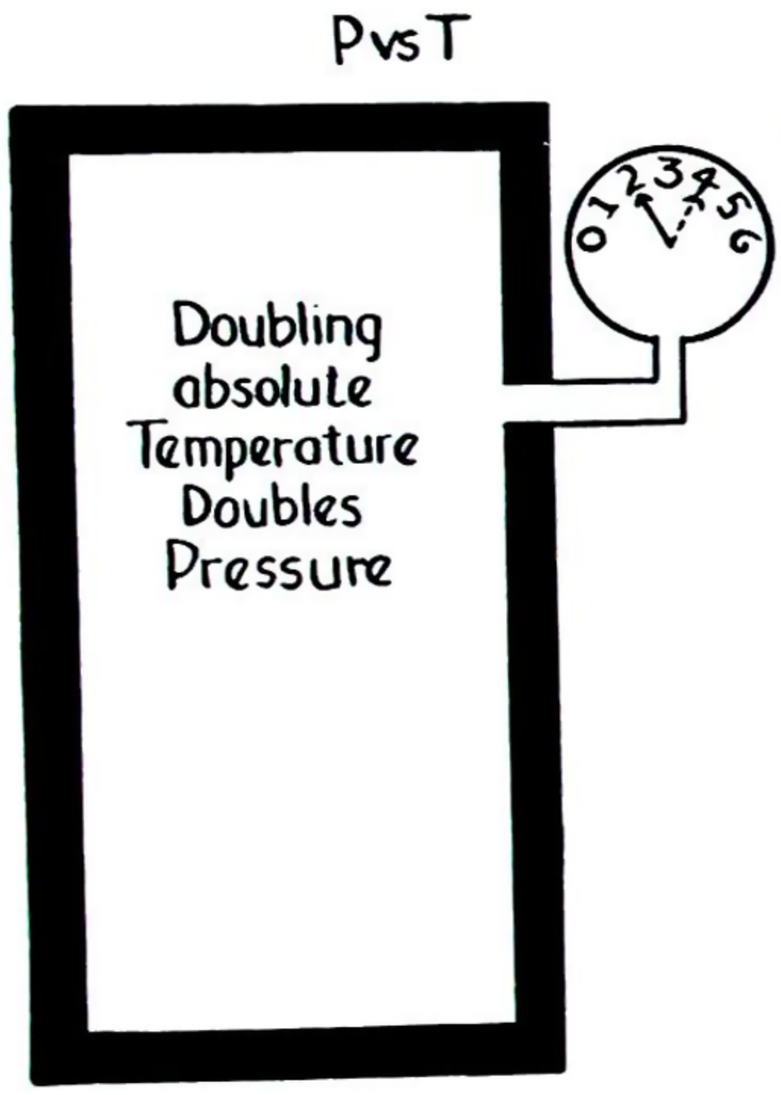


FIG. 72. Diagram illustrating the experimental facts expressed by the Gay-Lussac law.

> *If the volume is kept constant, the pressure on an enclosed mass of gas will vary directly with its absolute temperature.*

All of these facts represent things about gases that must be considered in any theory which is advanced as to the nature of matter.

The Phenomenon of Change of State

We have just been discussing the states in which matter is found. We know that our ordinary temperatures determine largely the particular state in which we commonly know any given substance. For example, we are quite familiar with water in all three states of solid, liquid and gas. Iron, on the other hand, we usually think of as being

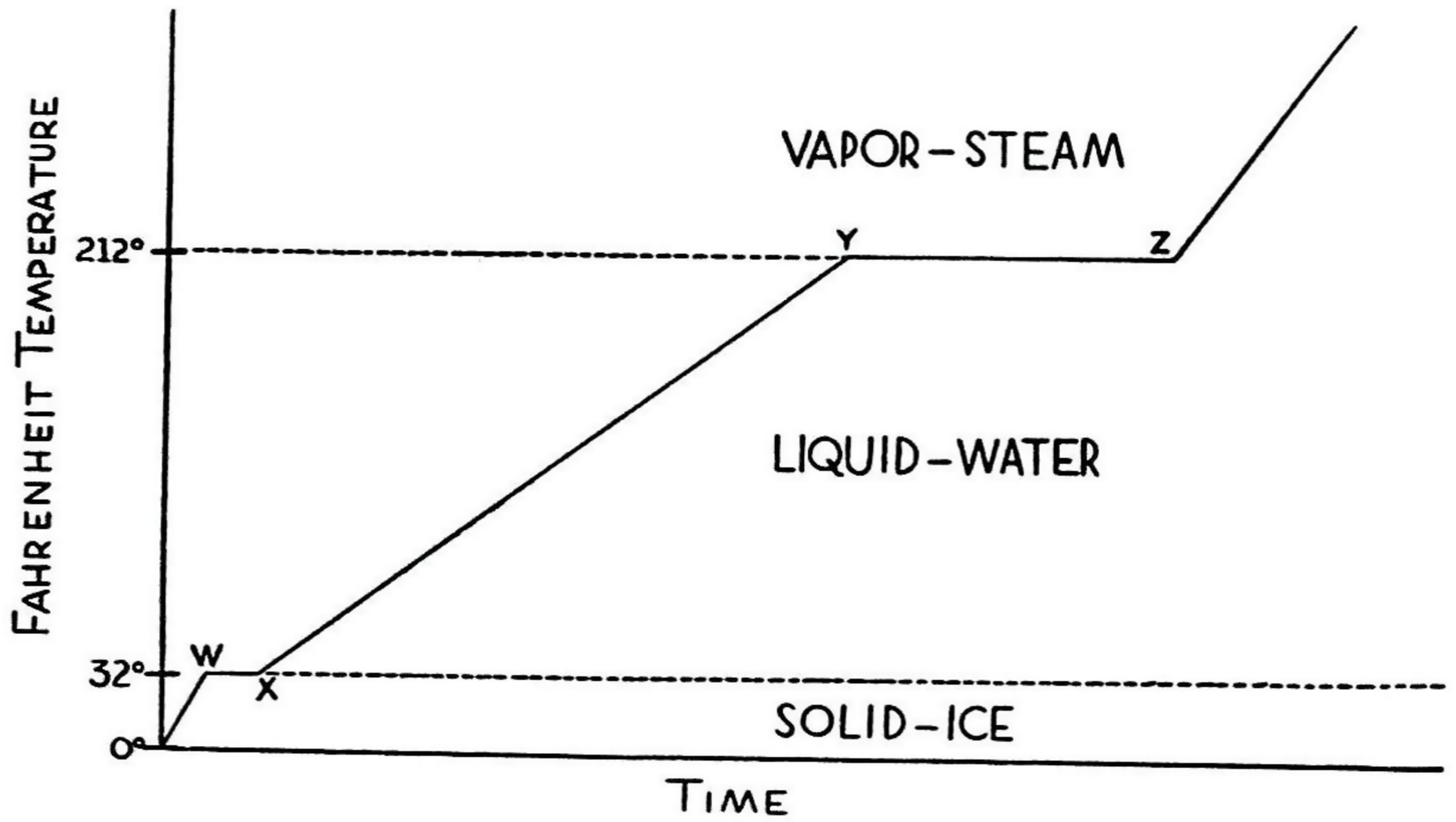


FIG. 73. Schematic representation of variation in temperature of H_2O during the time a quantity of it is heated from 0° F, to 212° F.

a solid although everyone knows it can be melted. We have always thought of carbon dioxide as a gas until recent years, when we have also become acquainted with it in its solid state as the refrigerant known by the somewhat misleading name of " dry ice." In fact, all known gases have been liquefied and solidified. Now, when we investigate this process of changing materials from one state to another, we observe a peculiar phenomenon.

As an example we will discuss a typical experiment of changing ice into water and then the water into steam. Assume that we have placed in a vessel a piece of ice at a temperature of 0° F. Then we place the vessel over a slow flame, and with the entire apparatus well insulated let the heat from the flame be absorbed by the ice. To check just what happens, we take the temperature of the H_2O at one minute

intervals throughout the experiment. The graph in Fig. 73 indicates schematically, although not to scale, the sort of results we might get in a typical procedure of this sort. Let us examine these data carefully. The ice, being considerably below freezing at the beginning, must be heated to 32° F. before it begins to melt. Consequently the first part of the graph, up to point *W*, indicates that the substance increased in temperature as it absorbed heat. This of course is the usual effect of the application of heat to a material. The part of the graph from *W* to *X*, however, shows that here was a period of time during which the heat, although it was being supplied at the same rate, caused no increase in temperature whatsoever. What is observed to occur during this time is the change of the solid into the liquid. At *W* it was all ice and just beginning to melt. Just at X it was completely melted and was all water. Evidently, then, this heat went into causing a change of state.

From X to Y, the usual effect of rise in temperature with absorption of heat is indicated, as the water is raised from 32° F. to 212° F. Beginning at 212° F., as point *Y* is reached, we again find heat being used to cause a change of state rather than to increase the temperature. Here the change is from a liquid to a vapor, or, in our example, from water to steam. Once the vaporizing process is complete at Z, more heat will again cause the temperature to rise.

These amounts of heat that are absorbed during changes of state are called *latent* or hidden heat, in somewhat the same sense that commencement orators speak of the latent possibilities in graduating classes. In the latter case they mean possibilities that may or may not make themselves known at a later time; in the former we mean heat that may come back out again. It will come back out, in fact, if the process is reversed because heat that is absorbed during the change from a solid to a liquid or a liquid to a gas will be emitted when the vapor condenses or the liquid changes to a solid.

From the standpoint of our search for a satisfactory theory of matter, this means that whether we like it or not, heat is forcing itself into the picture. In other words we cannot hope to get the complete story of the fundamental nature of matter unless we get some idea of what heat is, since it seems to play a leading role in switching matter about among its various states.

The first part of the next chapter concerns itself with the ques-

tion of the nature of heat, after which we will be in a position to discuss the theory that ties the two together.

Questions

Discussion

1. Explain the difference between a "continuous" and a "discontinuous" state of matter.

2. What effect did the teachings of Aristotle have on the development of an atomic theory of the nature of matter?

3. Name Aristotle's four elements and four attributes and relate them to each other.

4. Mention several ways in which our ideas about elements and atoms differ from those held by John Dalton.

5. Find out what element is represented by each of the following chemical symbols and state whether each is a solid, a liquid, or a gas at ordinary temperatures: Al; C; Au; He; I; Fe; Mg; Mn; Hg; Ne; K; Na; Ag; U; Zn.

6. Show how the periodic table aided in the discovery of new elements.

7. Distinguish between elements, compounds, and mixtures, and give examples of each.

8. Give a definition of a balanced chemical equation.

9. Look up some chemical equations, other than the ones given in this chapter, and find out all you can about the reactions described.

10. Find out how to balance the following equations:

1. $CaO + HCl \rightarrow CaCl_2 + H_2O$
2. $H_2O + Fe \rightarrow Fe_3O_4 + H_2$
3. $Cu + O_2 \rightarrow Cu_2O$
4. $NH_3 + O_2 \rightarrow NO + H_2O$

11. List several bases on which we commonly distinguish solids from liquids and liquids from gases.

12. Compare solids, liquids, and gases with one another as to compressibility and rates of diffusion.

13. State Boyle's law and the Gay-Lussac law and show what they mean when applied to a given mass of gas enclosed in a cylinder.

14. Draw a graph indicating the changes that occur in the state of H_2O when it is heated from 0° to 220° F.

15. Find out what the reaction is that is represented by the equation given in True-False Ques. 4.

16. From the discussion in this chapter can you find any explanation of the fact that open tubs of water placed in a fruit cellar will protect the fruit from freezing?

Multiple Choice

1. Helium has an atomic weight of 4; therefore (a) a cubic foot of it weighs 4 ounces; (b) 4 atoms of it weigh a gram; (c) 16 atoms of it weigh

as much as 4 atoms of oxygen; (d) a molecule of it contains 4 atoms; (e) an atom of it weighs as much as 4 atoms of oxygen.

2. Great resistance to compression is a property of (a) solids, liquids, and gases, (b) solids and liquids but not gases, (c) solids only, (d) liquids only, (e) solids and gases but not liquids.

3. Volume varies inversely with pressure in enclosed (a) solids, liquids, and gases, (b) solids and liquids, (c) solids and gases but not liquids, (d) liquids only, (e) gases only.

4. Aristotle believed that (a) there could be no such condition as a few fundamental elements out of which everything was made, (b) there were some 20 elements instead of the 92 we now know, (c) everything could be shown to be composed of hydrogen, (d) there were four elements namely, earth, air, fire, and water.

5. Which two of the following must occur before some perfume that is spilled in one corner of a room can be detected in the opposite corner? (a) Melting; (b) Diffusion; (c) Freezing; (d) Evaporation; (e) Compression.

True–False

1. The earliest suggestion made by anyone that matter might be atomic in nature was made about the time of Galileo.

2. The name of the Russian chemist, Mendeleef, is associated with the discovery of the difference between elements and compounds.

3. The procedure of sawing a large piece of wood into a number of smaller pieces represents a physical rather than a chemical change.

4. The following equation is balanced,

$$2NaCl + 2H_2O \rightarrow 2NaOH + Cl_2 + H_2$$

5. One of the characteristics that distinguish mixtures from compounds is the fact that only the former possess the property of definite composition.

6. The experimental facts with regard to diffusion would seem to suggest that in all solids, the molecules are packed tightly against each other.

7. In view of the definition of direct proportion given in this chapter, it is correct to say that for most people weight is directly proportional to age.

8. The name of Robert Boyle is associated with a law which states the relationship between pressure and volume of an enclosed gas at constant temperature.

9. It is possible for a substance to absorb heat without at the same time rising in temperature.

10. Carbon has an atomic weight of 12; therefore 16 atoms of carbon weigh the same as 12 atoms of oxygen.

Chapter 14

THE NATURE OF MATTER (Continued)

The Nature of Heat

The phenomenon discussed at the close of the preceding chapter indicates that we must make our peace with the nature of heat before we can reach a satisfactory picture of the nature of matter. Therefore, we shall digress in that direction during the first part of this chapter.

It may have occurred to the reader at some time or other to wonder just what happens to a vessel of water when it is heated, or just what this thing called heat is. This is another of the problems with which early man became concerned. He soon began to advance hypotheses about it. Before tracing the development of any of these theories it will be necessary to define a few terms and concepts.

Heat and Temperature

It is important that we have clearly in mind the relationship between the quantities, heat and temperature. Although they are often confused, a little thought will demonstrate that they are not at all identical. The amount of heat that would raise a cup of water from room temperature to boiling would increase the temperature of a large kettle of water only a few degrees. Rather than being simply different names for the same thing, the relationship may be thought of as one of cause and effect, as the above illustration suggests. When a substance absorbs heat the *usual* effect is an increase in temperature. When it gives off heat, it usually drops in temperature. Note the use of the word " usual," in view of the fact that we discovered in the last chapter that the effect of absorption or emission of heat is sometimes a change of state and not a change of temperature.

Since the more common effect is variation in temperature we find it necessary to be able to measure temperature. A device that performs this function is called a thermometer, and here we may mark up another first for Galileo. The principle of his first thermometer, or thermoscope as he called it, is shown in Fig. 74. The bulb of the flask contained an amount of air so that the water stood at some point in the neck, as *L*. Everyone is familiar with the fact that most substances expand as they rise in temperature. In this case, such a rise would cause the air to expand, forcing the water down. On the other hand, if the temperature dropped, the air would cool and contract, and the water would rise in the neck of the flask. The operation of Galileo's thermometer was, therefore, the reverse of the kind with which we are familiar. It is important to note that the entire operation depended on a change that occurred in matter as heat was absorbed. Similarly, the operation of every temperature measuring device depends on some property of matter that varies as heat is emitted or absorbed.



FIG. 74. Design of Galileo's thermoscope, or first thermometer.

Our common mercury and alcohol thermometers, like Galileo's thermoscope, operate as a result of the expansion of some substance. By knowing how the volumes of these materials vary with absorption or emission of heat, we are able to fix scales on them and get a numerical measurement of what we call temperature. The standard thermometer in use today, that is the one by which others are calibrated, is a gas thermometer. The gas usually used is hydrogen, although helium may be employed for very low temperatures and nitrogen for very high ones. The range of temperatures that can be covered with these gases extends from less than a degree above absolute zero up almost to 1,500° C., which is white heat. The value of this particular type of thermometer as a standard lies in the fact that it obeys accurately the gas laws over great temperature ranges.

Other types of temperature measuring devices are based on other heat-matter phenomena. It will be recalled that in Chapter 5 we discussed the thermocouple which is employed in measuring star tem-

peratures. The basis of its operation is the fact that, under certain conditions, temperature differences will cause an electric current to flow, the magnitude of the current depending on the difference in temperature. A device called the bolometer operates as a result of the experimental fact that the resistance which certain materials offer to the flow of electricity varies with the temperature of the substance. Furnace temperatures are sometimes determined with a series of objects called Seegar's cones. These cones have different melting points so that by finding which ones melt in the furnace, the temperature can be determined with a fair degree of accuracy.

With regard to temperature scales there are three which must be considered. Two of these are the more or less familiar Fahrenheit and Centigrade. The third is the Absolute or Kelvin scale which was mentioned in the preceding chapter in connection with the behavior of gases, and which will be taken up in detail later in this chapter.

It should be understood, that for any scale of temperature measurement to be satisfactory, two things must be true. First, it is necessary that two fixed points on it be clearly defined. Second, the two defining temperatures should be constant and easy to reproduce. For example the hottest day in summer and the coldest day in winter would not be satisfactory points to use as a basis of definition of a temperature scale. Neither would the temperatures of the North Pole and Sahara Desert be desirable. The freezing and boiling points of water, on the other hand, are quite satisfactory.

The two points chosen by Fahrenheit in fixing the scale which bears his name were the lowest temperature he could obtain with a mixture of salt and ice, and body temperature. He designated the former as zero and the latter as ninety-six. On any given thermometer, then, the space between the mercury levels at these two temperatures was divided into 96 equal parts or degrees, and divisions the same size extended above and below this range. This procedure made the freezing point of water 32 degrees and its boiling point, 212 degrees.

The Centigrade scale is used quite commonly in Europe by everyone and almost exclusively in scientific work everywhere. Its defining points are the temperatures of freezing and boiling water, the interval between being divided, as the name of the scale suggests, into

100 equal divisions. Thus the freezing point of water is 0 degrees C., and its boiling point 100 degrees C.

A comparison of the two scales is shown in Fig. 75. To transfer temperatures from one scale to the other is not a difficult procedure. As an example we may determine what a room temperature of 68° F. would be on the Centigrade scale. First, we note that on the Fahrenheit scale there are 212 — 32 = 180 degrees between the freezing and boiling points of water, while on the Centigrade scale the same range has only 100 degrees. Thus the number of Centigrade degrees in any given temperature range is 100/180 = 5/9 the number of the Fahrenheit kind; or the number of Fahrenheit steps is 9/5 as many as the number of Centigrade divisions. In changing 68° F. to Centigrade we will use the freezing point of water as the bridge from one scale to the other. To get to this bridge on the Fahrenheit side we must go down 68 — 32 = 36 steps or degrees. Crossing over we find ourselves at 0° C. To ascend on this side to a temperature opposite our starting point, we must climb 5/9 as many steps, that is 5/9 times 36, or 20. Since we started our climb at 0° C., our destination is 20° C.; in other words, 20° C. = 68° F. A similar procedure should make it possible for the reader to transfer any temperature from one scale to the other. For the benefit of those who prefer to use an equation, the following is offered,

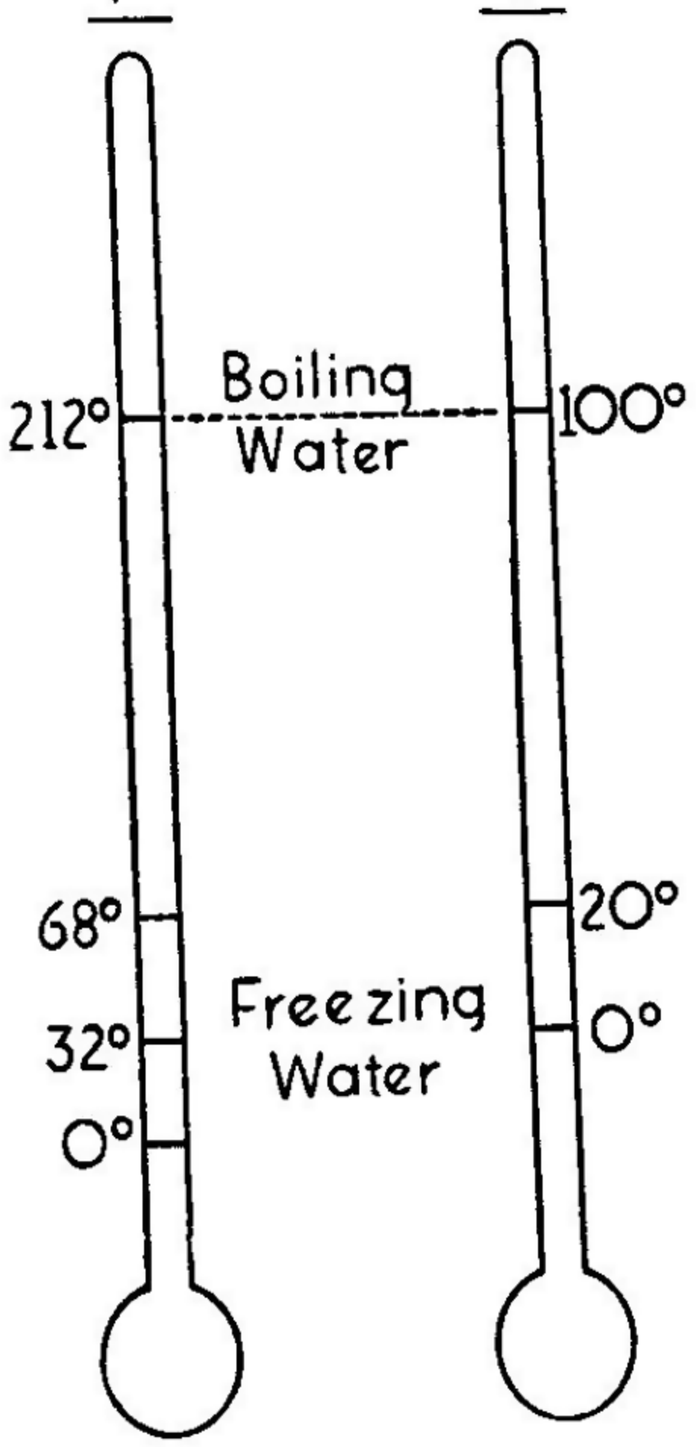


FIG. 75. A comparison of the Fahrenheit and Centigrade temperature scales.

$$5F - 9C = 160,$$

in which F and C stand for the Fahrenheit and Centigrade temperatures respectively. By substituting the value of either, the other can be calculated. Temperatures below zero on either scale are used as negative.

The Absolute scale will be related to the two just discussed after its physical significance has been developed.

Having considered the measurement of temperature, we are now in a position to define two units of heat. These are the calorie and the British thermal unit. The first is a metric unit and may be defined as the average amount of heat required to raise the temperature of 1 gram of water through 1 degree Centigrade. In an analogous fashion, the British thermal unit, commonly called the B.t.u. is the average amount of heat necessary to raise the temperature of a pound of water 1 degree Fahrenheit.

Some Experimental Facts Relating Heat and Matter

Here again, we will attempt to proceed in accordance with the scientific method by first discussing several typical phenomena relating heat and matter, and then developing a theory that will make them seem reasonable.

Absorption of Heat. — Since this phenomenon has already been given some attention we will simply recall to the reader's mind a few of the facts and mention several examples. We found that when a substance absorbs heat, one of two things happens. Either it increases in temperature or it changes state in the direction of solid to liquid, or liquid to gas. The reverse is true when heat is given off. A satisfactory picture of the nature of heat will first have to give some physical significance to this concept of temperature, and second, account for the heat that goes into changing temperature and changing state.

Expansion with Rise in Temperature. — It is common knowledge that most substances expand as temperature rises and contract as it falls. One of the most familiar examples is the thermometer, which we have just been considering. Many other illustrations could be cited. Steel bridges lengthen in summer and shorten in winter. A steel structure like the George Washington bridge across the Hudson River, which has a span of some 3,500 feet, may be several feet longer at noon on July 4 than it is at midnight on December 31. This expansion is usually allowed for in one of two ways. Either one end of the bridge is mounted on a roller instead of being rigid with the abutments, or the structure is built with an expansion gap in the center. Similarly, expansion joints are inserted in concrete roads, and railway rails are laid with small gaps between adjoining ends. Fig. 76 shows a simple thermostat whose operation is based on this principle. Strips

of two metals, *A* and *B* are riveted together. They are so chosen that *A* expands more rapidly than *B* with increase in temperature. If the temperature rises, the double bar will bend to the left as shown and will eventually make contact, closing the circuit and permitting the bell to ring. In this way variation in temperature is used to open and close an electric circuit.

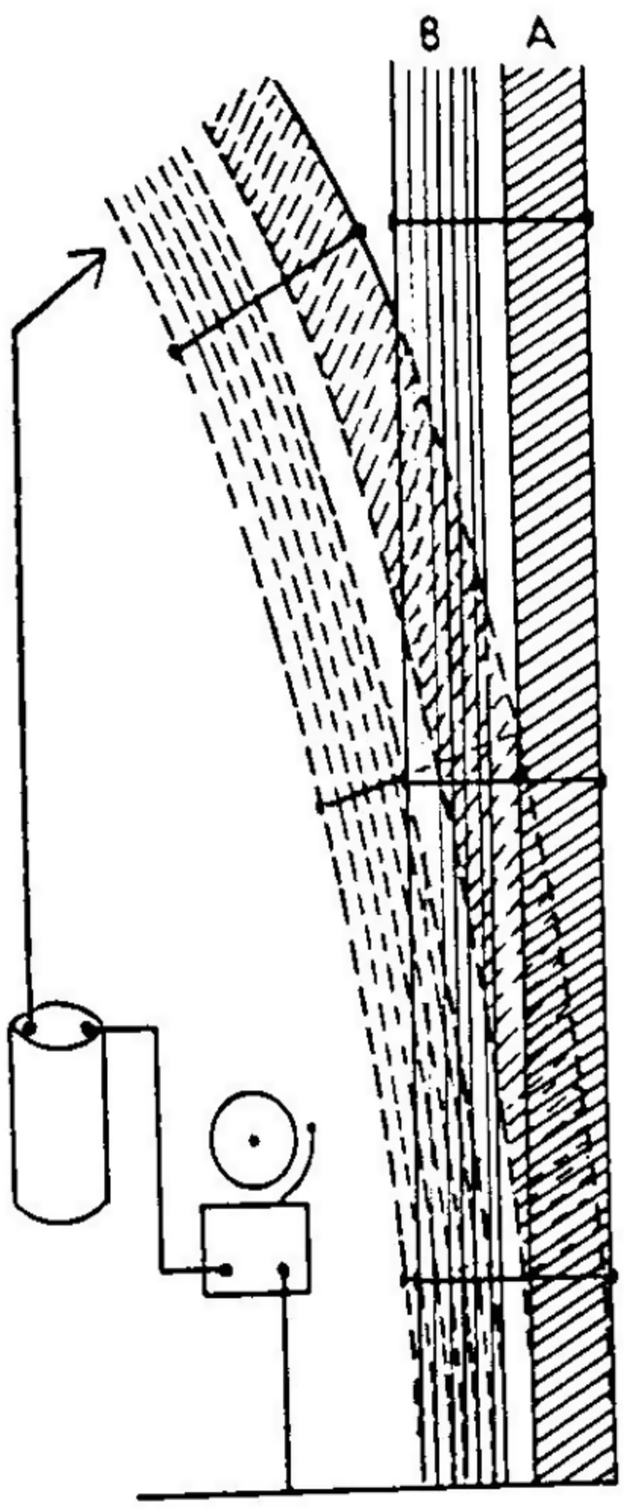


FIG. 76. A simple thermostat placed in an electric circuit with a bell.

Any satisfactory theory of the nature of heat must give a reasonable explanation of this phenomenon of expansion.

Heat Transfer. — There are three ways in which heat may be transferred from one place to another. The first of these may be illustrated by the manner in which the handle of a coffee spoon becomes heated. Although only the bowl of the spoon is in contact with the hot liquid, the temperature of the handle soon rises also. We say this occurs as the result of *conduction* of heat along the material (Fig. 77). Materials differ in the rapidity with which they will conduct heat. For uses such as insulation in houses and refrigerators, or handles of cooking utensils, poor conductors like wood, cork, or asbestos are desirable. Good conductors, such as the metals, are employed in cooking vessels, steam and automobile radiators, and similar devices in which it is desired to get heat transferred quickly from one place to another.

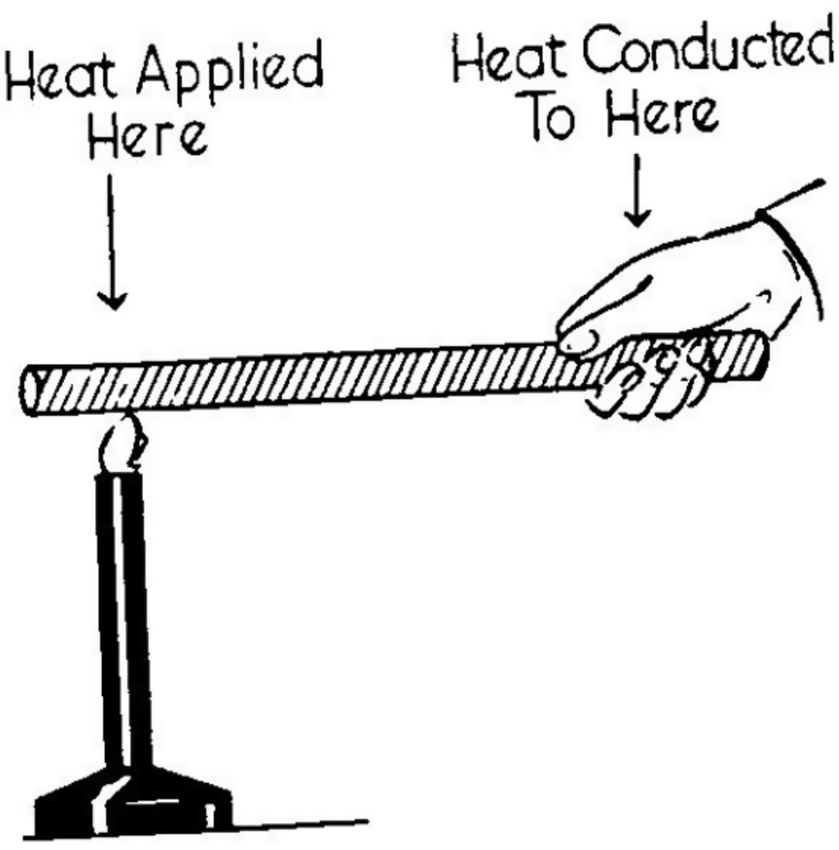


FIG. 77. Heat transfer by conduction.

The second method finds its application in liquids and gases, and may be illustrated by the manner in which water in a vessel on the stove becomes heated. The water at the bottom is heated first, after the heat has been conducted through the bottom of the container. The heated water expands, becoming

lighter per unit volume than the surrounding liquid, and so rises to the top. Cold water flows in to fill the vacant space and in turn becomes heated. The process continues, setting up in the liquid currents which are called convection currents, shown in Fig. 78. This method of heat transfer is called *convection*. These currents are the same sort as those discussed in connection with a previous chapter on the causes of weather. It will be recalled that the heated air over the equator rises, causing winds to blow toward the equator from the north and south. Note that in this method of transfer the actual portions of the substance being heated move out of the way, allowing other parts to move in and be heated, whereas in the case of conduction the heat is more or less handed along from each part to the next.

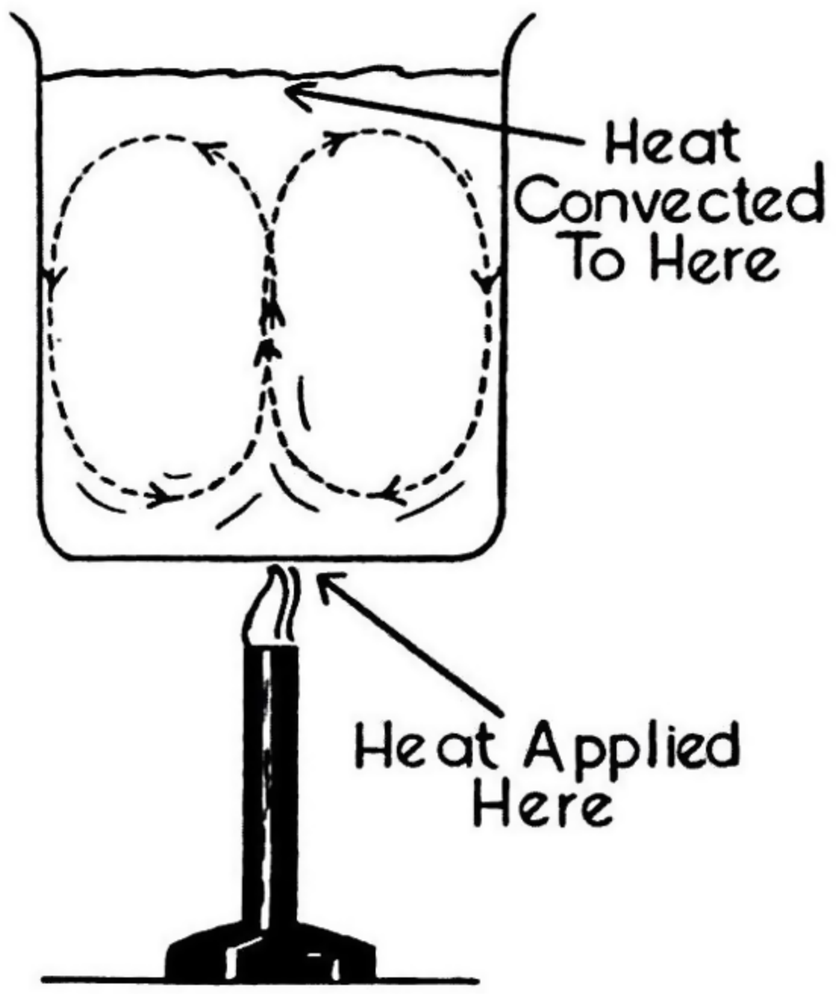


FIG. 78. Heat transfer by convection.

One may illustrate these two methods by two ways in which food could be served to a large group. Those eating might all be seated and the plates passed from the kitchen to a waiter, to the person at the end of the table and along to the rest of the diners, with no one having to move from his place. Such a procedure would be analogous to conduction. Or, the food could all be placed on one table and those eating file past, each gathering up his own meal, as in a cafeteria, in which case the meal would be served in a manner comparable to convection.

The third method of heat transfer represents, in a sense, an invention in order to provide for cases which fit neither of the above. In fact that most important of all examples of heat transfer, cannot occur by either conduction or convection. This is the transfer of heat from the sun to the earth. Both conduction and convection require the presence of a material substance of some sort, while between the earth and the sun there exists a high vacuum. Similarly, most of the heat that one gets from an open fireplace must travel by some other means than conduction or convection, since air is a very poor conductor and

the convection currents in it are toward rather than away from the fireplace. The fact that one may feel the warmth of sunlight coming through a window although the glass itself remains cold, also indicates the existence of a third method of heat transfer. In all of these cases the heat is said to travel by *radiation*. The discussion of radiation in some detail will be saved for a later chapter. For our purposes here it will be sufficient to say that it still presents certain unsolved problems which are principally connected with the medium in which the heat waves travel. Some of the assumptions made in this connection are excellent examples of the human tendency to invent a solution to a troublesome problem and feel very happy about it, ignoring more or less completely the new problem of accounting for the invention.

These methods of heat transfer also represent experimental phenomena which any good theory of heat and matter must substantiate.

Before considering our present theory of the nature of heat and matter, we will look briefly at an early theory of heat which was accepted for many years.

The Caloric Theory of Heat

It had long been recognized that there were two possibilities with regard to the nature of heat. Either it was an actual material substance, or it was a form of energy. A number of early investigators including Newton and Cavendish favored the latter notion. That is, they thought of heat as being due to the motion of the molecular particles of the substance involved. However, they were unable to devise a complete working theory on this basis, and science in general adopted the other alternative. In terms of this theory, heat was pictured as a fluid which was given the name " caloric." It is worth noting that early theories concerning the nature of electricity also passed through a similar " fluid " stage.

The fundamental assumptions of the caloric theory of heat were as follows:

1. Heat is a fluid called caloric.
2. Caloric raises the temperature of any substance to which it is added, and lowers the temperature of any material from which it is removed.

3. Caloric flows naturally from substances of higher temperatures to those of lower.

One of the earmarks of a good scientific theory is that a few original assumptions account for a large number of phenomena. In the above group there are one definition and two assumptions, each of the latter accounting for one experimental fact. In order to take care of new experimental facts, it soon became necessary to make additional assumptions. For example, in order to make expansion with rise in temperature seem reasonable, it was assumed that,

4. Caloric is self-repellent.

Thus, the more caloric a substance acquired, the greater would be the total repulsion and the more the material would expand.

Then it was recalled that every fluid known to man had weight. This, in the light of the second assumption, made it appear that an object should get heavier as its temperature rose. Experimentally, however, this was found not to be true. So another assumption was made, to the effect that,

5. Caloric is a weightless fluid.

Although this assumption seems quite ridiculous in the light of our present knowledge, it may be mentioned that the idea of an ether in which heat and light waves travel came into existence in somewhat the same manner.

The caloric theory was placed in the position of acquiring a new basic assumption every time a new experimental fact made its appearance, which is a precarious situation for any scientific theory.

The actual overthrow of this picture of heat resulted from the experimental work of a number of different men. Among the first of these was Benjamin Thompson (1753–1814), better known as Count Rumford. Thompson was a native of Massachusetts who was forced to leave this country during the Revolutionary War because of his pro-British sympathies. He accepted a commission in the English army and later saw military service in Bavaria, where he received his title. His scientific activities were numerous and included serving as one of the founders of the Royal Institution in London about 1800.

His work of particular interest here began while he was engaged in manufacturing cannon in Munich. He became curious as to the origin of the large amounts of heat that were generated during the boring of the cannon, and presented the supporters of the caloric theory with the problem of explaining it. Since everything around it had a lower temperature than the cannon, it was difficult to see, in the light of the third basic assumption, where the caloric could come from. However, the believers in caloric advanced the very ingenious suggestion that a pound of powdered iron contained less heat than the same mass of bulk iron at the same temperature; or to state the point more technically, that the specific heat of powdered iron was less than that of bulk iron. Thus, they argued that as the iron was powdered during the boring process, the excess heat would be given up. Count Rumford tested this hypothesis by changing to a blunt boring tool and found that much more heat was generated, although far less powdered iron was produced.

The experiments of Sir Humphry Davy added to the troubles of the caloric theorists. Davy devised an apparatus by which two pieces of ice were rubbed together mechanically in a vacuum. He found, as we would expect, that the ice melted. In terms of the caloric theory the problem again was one of accounting for the source of the heat, since the ice was cold and was separated from its surroundings by a vacuum.

The final and complete abandonment of the theory resulted from the experiments of James Prescott Joule (1818–1889). The work of Rumford and Davy, although qualitative, had indicated the likelihood that the generation of heat was closely related to mechanical work or energy. During the decade, 1840–1850, Joule performed a series of experiments designed to test this possibility in a quantitative manner. The design of the apparatus used is shown in Fig. 79. The vessel used bore some resemblance to an old-fashioned ice cream freezer in that it contained a paddle arrangement (P) the vanes of which moved between other vanes (V) which were fixed to the vessel wall. A rapid rotation of the paddles churned the water in the cylinder and caused its temperature to rise. This rise could be noted on the thermometer. Above the cylinder were two weights (W) attached to cords wound around a pulley on the shaft to which the paddle vanes

were fixed. The weights then, when allowed to fall, caused the rotation of the paddle wheel. From the weights and their distance of fall it was possible to calculate the mechanical work done; while the weight of water in the cylinder and its increase in temperature were sufficient data for the determination of the amount of heat generated. The entire apparatus was insulated against outside effects.

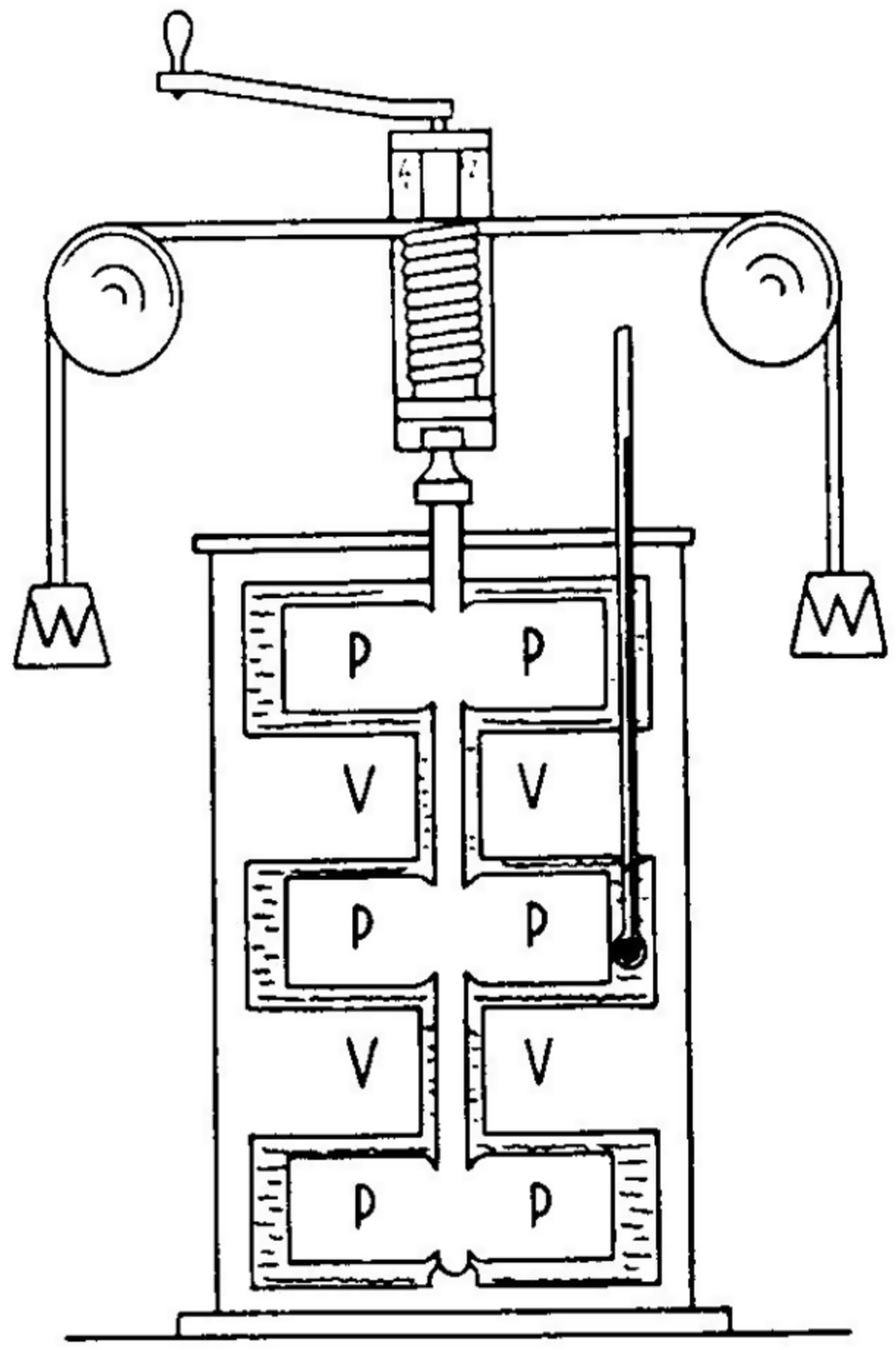


FIG. 79. Cut-away diagram of type of apparatus used by Joule in determining the relationship between mechanical work and heat.

Joule reasoned that if heat were a form of energy rather than a fluid, a given amount of work should always produce the same amount of heat. He performed the experiment time after time, varying the amount of water in the vessel, the size of the weights, and the distance of fall. In every case he found, within the limits of experimental error, the same numerical relationship between mechanical work and heat. This result may be expressed by saying that 778 foot-pounds of mechanical work are required to generate 1 B.t.u. of heat. The B.t.u., or British thermal unit, of heat has already been defined. One foot-pound of work may be thought of as the work that is done when a weight of 1 pound is lifted through a vertical height of 1 foot. Thus to heat 1 pound of water 1° F. takes as much energy as that required to lift a 778-pound weight 1 foot, or a 1-pound weight 778 feet. Let us emphasize again that the important thing, from the standpoint of the nature of heat, is the fact that substantially the same numerical relation was obtained in all of Joule's experiments, and in all that have been performed since that time. Another point to note is that here again the conclusive experiment was a quantitative one, one in which specific numerical results were obtained.

The Kinetic Theory of Heat and Matter

Basic Assumptions

The fundamental assumptions upon which the kinetic theory is based may be stated as follows:

I. Matter is made up of molecules which in turn are composed of atoms, the relationship between these being as described in the preceding chapter.
II. Molecules are in constant motion with velocities which increase with temperature.
III. Molecules attract each other with forces which obey a law similar in form to the law of gravitation.

Experimental Facts in the Light of the Kinetic Theory

Having set up the basic postulates of the kinetic theory, let us return to the experimental facts presented in the preceding chapter and examine their reasonableness in the light of this theory.

Shape and Rigidity. — The second of the above assumptions can be true only if matter is to some extent porous, that is, if the actual particles of materials are separated by empty spaces. By combining this idea with the third assumption, we arrive at a consistent picture of the nature of the three states of matter. It will be recalled that in Newton's law of gravitation, the force decreases with the square of the separation. The situation is similar for attraction between molecules. In solids then, we would think of the molecules as being comparatively close together, close enough that the forces of attraction are capable of holding the material in the definite shape that is characteristic of this state. A somewhat greater average separation of molecules might be expected to produce the condition that exists in liquids, namely, attractive forces large enough to result in maintenance of a free surface but not sufficient to give definite shape. Still greater spaces between molecules could reduce the forces of attraction to the point where they would become negligible and the molecules would move freely throughout all the available space, as happens in the case of gases.

To summarize then, the kinetic theory would distinguish the

three states of matter on the basis of the average molecular separation that exists. Such a picture carries with it the implication that change of state could be accomplished by varying this average separation.

Compressibility. — It will be recalled that gases are found to be readily compressible, liquids almost incompressible, and solids essentially incompressible. This sequence seems to agree well, qualitatively at least, with the above picture. The comparatively large molecular separations hypothesized for gases should make them the easiest to compress, while solids would be expected to offer the greatest resistance to being forced into a smaller volume.

Diffusion. — The assumed molecular separations of the three states again make it seem entirely reasonable that the order of ease and rapidity of diffusion from greatest to least should be gases, liquids, solids. The evidence indicates that in gases at normal pressures the sizes of the molecules bear something of the same ratio to their separations that the diameters of the stars do to their distances apart. Also molecular velocities in gases may run as high as a mile per second. Consequently, diffusion of gases would be expected to proceed at a high rate.

Even in solids the empty space considerably exceeds that actually occupied by particles. Dr. W. F. G. Swann has calculated that if all the empty space were to be squeezed out of the much-discussed average man, his size would diminish to about that of an average flea, although his weight would remain unchanged. So that, instead of being surprised at the slight diffusion found in solids, one might be more likely to wonder why it is not much greater. One might think that out of all the automobile collisions that occur, the molecules of one car might occasionally sift in between those of the other and so avoid any damage, or that once in a while a fist would pass harmlessly through a jaw. The answer is that such a probability does exist, one in fact that is definitely capable of being calculated, for any given case. Dr. W. F. G. Swann cites the following example: [1]

"You may ask, 'Is there not a definite possibility that sometime as I start to fall through a chair, all the molecules of my body will miss the molecules of the chair so that I will fall right through?' Yes, that may happen if I live long enough. . . . I find that for a chair one inch thick I shall probably fall through once in N times, where N is such a large number that if an angel

[1] Swann, W. F. G., *The Architecture of the Universe*, p. 48. New York: Macmillan, 1934.

flying with the speed of light (186,000 miles per second) were to write it down in his flight, and if he were to write small enough that his figures were as tiny as electrons, and so fast that the successive zeros in the long number touched each other, he would have to travel for a number of years which is so large that it would require more than 60,000 figures to write it down. I shall not worry about the risk."

One of the reasons for this small probability is the rapid motion of the molecules. For example, in a bicycle wheel the empty space considerably exceeds that actually occupied by spokes. Yet when the wheel is revolving rapidly it offers a very solid-like resistance to penetration by a stick or by the hand. It is also true that while intermolecular space is large compared to the sizes of the molecules themselves, yet it is extremely small when compared with ordinary dimensions, even of a relatively small object. As an illustration, suppose you should find some evening that, while you were out in the car, someone had packed your garage completely full from floor to ceiling with woven wire wastebaskets. From the standpoint of driving your automobile inside, the garage would present most of the properties of a solid, although to a housefly or mosquito it would seem relatively empty. In the same way, a brick wall might offer little opposition to the entrance of a single molecule but make penetration by another brick extremely improbable.

Evaporation. — The experimental facts involved here are, first, that liquids evaporate, and second, that the rate of evaporation increases with the temperature. Evaporation is simply a more or less gradual change from the liquid state to the vapor. Consequently, assuming all the molecules to be in motion it is reasonable to suppose that some would strike the surface with sufficient velocity to escape into the atmosphere.

Since the observed rates of evaporation vary with temperature, we must see what sort of significance the concept of temperature has in the kinetic theory. Our discussion of the nature of heat concluded with the evidence that heat is a form of energy. Therefore, if a substance increases in temperature as it absorbs heat, that temperature must be a measure of its energy. One form of energy which any object may possess is that due to motion, called Kinetic Energy. A plausible representation of the absorption of heat energy, then, is sim-

ply the speeding up of the molecules of the substance, or the increasing of their kinetic energy. On this basis, temperature is simply a measure of the average velocity of the molecules of the material in question.

Now let us return to the variation in rate of evaporation with temperature. Increased temperature, according to the above picture, should provide an increased number of molecules with velocities high enough to permit them to escape from the liquid. In other words, rising temperature would be expected to result in greater rapidity of evaporation.

Gas Laws. — The first of the two gas laws previously mentioned was Boyle's law which involves the relationship between pressure and volume of an enclosed gas. The picture of a gas as consisting of molecules in rapid motion would indicate pressure to be the total force exerted by these millions of moving particles as they strike the sides of the container. The situation would be analogous to that of a sand blast which exerts a fairly uniform total pressure as the result of millions of tiny blows by individual sand particles.

If a certain amount of gas is confined in the cylinder as shown back in Fig. 71, it will have a certain pressure and temperature. If the piston is pushed down until the gas is limited to just half the original volume, and if the temperature, which is a measure of the average velocity of the gas molecules, is held constant, then it would only be reasonable to expect that the number of blows against the sides of the cylinder would be doubled. In other words, the pressure would become twice as great. Or, if the volume were reduced to one-third, the number of blows per second and the pressure would be tripled. It should be noted that keeping the temperature constant means that the molecules are not hitting any harder on the average.

The other gas law concerned the variation of pressure of an enclosed gas with changes in temperature when the volume remains the same. On the kinetic theory we have said that increase in temperature means increase in average velocity of the molecules. And if they move faster, they should hit harder against the sides of the container. That is to say, the pressure should become greater.

Heat Absorption. — Absorption of heat with rise in temperature has already been mentioned in connection with evaporation. It will

be recalled that it was pictured as speeding up the molecules. It may be noted that this conception is not troubled by the difficulty which beset the caloric theory. No change in weight would be expected as temperature increased, and the source of the heat in Rumford's and Davy's experiments was simply friction caused by mechanical work.

The alternative result of heat absorption, namely, change of state, can be pictured just as reasonably. Remember that molecules attract each other. Therefore it takes work to pull them farther apart. But pulling them farther apart is just what has to be done to effect a change from solid to liquid or from liquid to gas. In other words, the heat that we called " latent heat " is expended in increasing molecular separation. If this is true, then the latent heat should be given up again when the material changes state in the direction of gas to liquid or liquid to solid. Experimentally, this is the case, and is the reason that steam at 212° F. will burn one much more severely than will water at the same temperature, other things being equal. In the case of the steam, the skin is subjected not only to the high temperature of 212° F., but also to the latent heat which is emitted as the steam condenses.

Expansion. — It seems reasonable to expect that as average velocities of molecules increase, that is, as temperature rises, the boundaries of the material will be extended.

Heat Transfer. — With regard to conduction, let us recall the former example of the spoon in the hot coffee. In line with the kinetic theory, the average velocities of the molecules in the bowl of the spoon would increase first, because of their direct contact with the heated liquid. More rapid vibratory motion of these molecules would cause them to strike adjacent ones with greater force and set them in more rapid motion. This would continue throughout the length of the spoon until finally the molecules at the end of the handle would be traveling with a higher average velocity, that is, would be at a higher temperature.

Once expansion with rise in temperature becomes reasonable, then heat transfer by convection offers no difficulties, since that is the phenomenon upon which it depends.

The Absolute Temperature Scale. — It will be recalled that the zero points on both the Fahrenheit and Centigrade scales of tempera-

ture were chosen quite arbitrarily and without any physical significance in mind. We shall see that the zero point of the third scale, the Absolute, has a very definite meaning in terms of the kinetic theory.

In approaching this scale of temperature, let us consider a hypothetical experiment. Suppose the amount of gas in the cylinder in Fig. 80 is so chosen that, at a temperature of 0° C., its pressure multiplied by its volume will equal the number 273. These are the starting conditions of the experiment and could be achieved by varying the amount of gas that was put into the cylinder. For our purpose here the units involved need not concern us. Now given these starting conditions, *it is an experimental fact*, that if the temperature is lowered one degree to —1° C., the new product of pressure and volume will be 272. At —2° C., the product is 271. It is obvious that if this relationship should continue and the material remain gaseous, at —273° C. the product of pressure and volume would be zero. Now if either one of two quantities is zero, then their product is zero, so we must inquire as to the physical significance of zero pressure or zero volume according to the kinetic theory.

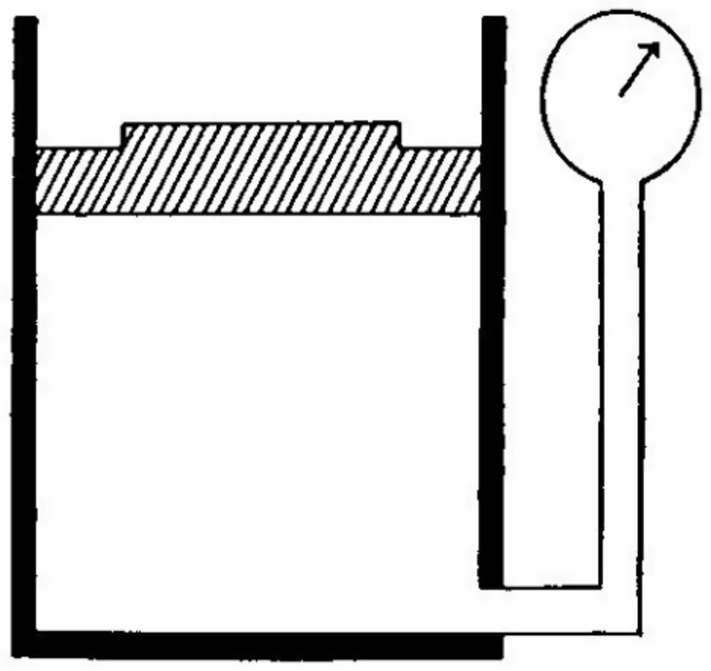


FIG. 80. Diagram illustrating the reason for Absolute Zero being just where it is.

It would seem that the only way in which volume could become zero would be for matter to be annihilated. But this is contrary to the evidence that we have for the essential indestructibility of matter. For the pressure of a gas to become zero, on the other hand, would merely mean that molecules were no longer striking the sides of the container. This would be true if they stopped moving, a perfectly reasonable condition in the light of the kinetic theory conception of temperature. This implies that —273° C. is the lowest temperature that would be possible since objects that have stopped moving cannot be further slowed down. Consequently the name of Absolute Zero is certainly a fitting one.

We see then, that the kinetic theory provides a temperature scale, the zero point of which has definite physical significance. As the comparative diagram for all three scales indicates (Fig. 81), the degrees on the Absolute scale are the same size as those on the Centigrade, while its zero is 273 degrees below zero Centigrade.

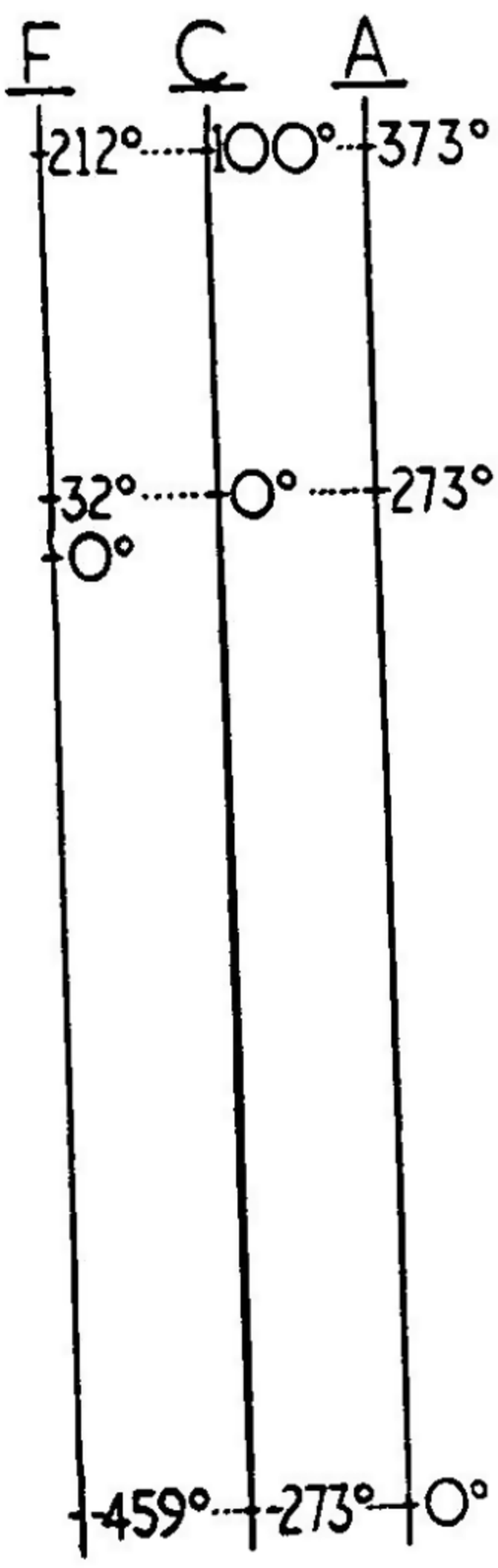


FIG. 81. A comparison of the Fahrenheit, Centigrade, and Absolute temperature scales.

The lowest temperatures actually obtained in the laboratory have been within less than a degree of this lowest possible point.

Brownian Movements. — One other evidence for the kinetic theory will be mentioned briefly. It concerns a phenomenon first observed in 1827 by Robert Brown, an English botanist, and called "Brownian movements." While examining, under a high power microscope, mushroom spores which were floating on water, he noticed that they never came to rest but were continually in motion in haphazard, zig-zag paths of the sort shown in Fig. 82. The kinetic theory offers the only explanation of this striking occurrence which seems to fit the observed facts. The spores behave exactly as we would expect tiny particles to behave if they were being continually bombarded from all sides by rapidly moving molecules of water. Also, we would expect the degree of agitation to increase the smaller the particle. This likewise is true experimentally. Dust and smoke particles in the air, microscopic bits of metal in a liquid exhibit the same phenomenon. While this evidence seems fairly conclusive and somewhat more direct than some of the others that we have discussed, it should be noted that it cannot be called absolutely direct proof of the kinetic theory. In the Brownian movements, one does not see individual molecules; one sees only particles which are aggregations of mol-



FIG. 82. Diagram illustrating Brownian movements.

ecules moving about in a manner that we are able to explain only on the basis of the kinetic theory.

With this we will conclude our discussion of the kinetic theory of heat and matter. The question has been considered in some detail for two reasons. In the first place, it is important that the reader have some understanding of the theory itself in view of the fact that it is generally accepted at the present time. In this connection, the reader should be reminded that so far we have been interested in the problem of the nature of matter only from bulk matter down to atoms and molecules. A little later we will concern ourselves briefly with the inside of the atom. In the second place, the theory has been considered as another example of the way in which scientific hypotheses come into being and grow, being constantly tested in the light of new experimental facts.

The next several chapters will concern a number of specific materials about which man has had to learn in order to bring about his present environment.

Questions

Discussion

1. Distinguish between heat and temperature.
2. Draw a diagram of Galileo's original thermoscope and show how its operation differed from that of a modern mercury thermometer.
3. Name and discuss the operation of two temperature-measuring devices which are based on some other principle than the expansion of a material with rise in temperature.
4. Three temperature scales were discussed in this chapter; within what range must a temperature be that is positive on two of the scales and negative on the third?
5. Express normal body temperature on all three scales.
6. Define the calorie and the British thermal unit.
7. How many calories are required to equal one B.t.u.?
8. Under what sort of conditions is it possible for a material to absorb heat and still not increase in temperature?
9. Draw a diagram of a simple type of thermostat and explain just how it is possible to use temperature changes to turn an electric current on and off.
10. Mention three methods of heat transfer and discuss the two taken up in some detail in this chapter.
11. State the basic assumptions of the caloric theory and discuss the experimental evidence which led to its abandonment.

12. State the basic assumptions of the kinetic theory of heat and matter.

13. Show how the kinetic theory accounts for the experimental facts which could not be explained by the caloric theory.

14. Discuss the relationship between the kinetic theory and the experimental facts in connection with the following:

a. Compressibility of matter in the various states.

b. Diffusion of matter in the various states.

c. Boyle's law.

d. Absorption of heat with and without rise in temperature.

e. Brownian movements.

15. Under conditions such that the product of pressure times volume for a gas equaled zero, why should one assume the pressure to be zero rather than the volume?

16. What is the physical significance of the 273 that appears in a definition of the Absolute scale of temperature?

Multiple Choice

1. If, of any two good thermometers hanging side by side, one reads above zero and the other below, which of the following (one or more) may be true? (a) The temperature must be above the freezing point of water. (b) The two instruments could be Absolute and Fahrenheit respectively. (c) The temperature must be below zero Fahrenheit. (d) The instruments could be any two of the three types. (e) The temperature must be lower than 273° Absolute.

2. An explanation of the fact that a rug feels warmer than the wooden floor on a cold morning, although both may be at the same temperature, would involve principally the phenomenon of (a) heat of fusion, (b) heat of vaporization, (c) heat conduction, (d) heat convection, (e) diffusion.

3. A given amount of gas in a closed container is heated from room temperature to 100° F.; disregarding possible expansion of the container, according to the kinetic theory (a) the gas has increased in weight, (b) the individual molecules of the gas have expanded, (c) the average distance between molecules has increased, (d) the average velocity of the molecules has increased, (e) the pressure on the sides of the container has decreased.

4. Large vessels of water placed in a fruit cellar on a cold night might be expected to (a) help keep the fruit from freezing, (b) have no effect on the freezing of the fruit, (c) make the fruit freeze more easily.

5. List all of the following discoveries which, if made, appear to be contradictory to the kinetic theory. (a) A gas that liquefies at 300° below zero Centigrade. (b) Two solids that would diffuse into each other. (c) A greater force of attraction between two molecules of steam at 212° F. than between two of water at the same temperature. (d) A material in which

all molecular motion stopped at 0° F. (e) A solid which absorbed heat as it changed to a liquid.

True–False

1. Water at a temperature of 400° Absolute would ordinarily be steam

2. The heat that a substance absorbs in changing from a liquid to a gas is believed to be used in increasing the average separation between molecules

3. If the water in the bottom of a vessel became heated due to the application of a blowtorch to the surface, the principal method of heat transfer would be conduction.

4. If the product of pressure and volume of a certain mass of enclosed gas equals 819 at 0° C., it would equal 759 at —20° C.

5. In the phenomenon of Brownian movements, the observer actually sees molecules darting about in a zig-zag fashion.

6. The vacuum between the walls of a thermos bottle is primarily for the purpose of preventing heat transfer by radiation.

7. The fact that the rate of evaporation of a liquid increases with rise in temperature constitutes one of the chief bits of evidence contradictory to the kinetic theory.

8. If the temperature of an enclosed mass of gas is kept constant and the pressure tending to compress it is doubled, the volume will be reduced to one-fourth its original value.

9. Caloric was assumed to have negative weight in order to account for the fact that heating a substance does not make it heavier.

10. A substance at Absolute zero could be said to contain no heat energy

hapter 15

IR AND WATER

)F all the substances in the world, these two unquestionably rank rst in importance to the human race. Man can live longer without ny other material than he can without air, while second in order of eed for survival comes water. It is natural that they should be the rst specific substances to be considered in the " Man and Materials " ortion of this book.

Air

The atmosphere in its relation to the earth was discussed in Chaper 7. Here we will be concerned with air as a material substance and is it is related to man, taking up first its properties as a gas and then is a liquid. Finally, we shall consider briefly its various constituents.

Properties as a Gas

The ancient world, it will be recalled, thought of air as one of the four elements. Today we recognize that it is neither an element nor a compound but a mixture. All of its fixed constituents but one are elements and that one is a compound. The percentage composition of air is given back on page 116.

One of the most important properties of air, as of all other substances, is that it has weight. Man has not always recognized this but thought of air for a long time as being weightless because it seemed to offer almost no resistance to his movements. But experiments have shown that just as a bucket full of water weighs more than an empty bucket, so a flask full of air weighs more than one that has been evacuated. At ordinary temperatures and pressures air weighs about 1.2 ounces per cubic foot.

Anything that has weight exerts pressure upon the material on

which it rests. Air, which we have found to be piled up to a height of many miles, pushes down on every square inch of surface at sea level with an average force of about 14.7 pounds. This is true of the earth's surface as well as of that of any object on the earth. The reader's own body, for example, is constantly being subjected to this same pressure, which amounts to more than a ton for every square foot of body surface.

The measurement of this atmospheric pressure is of considerable importance. We have already seen the part it plays in weather forecasting in our earlier discussion of that subject. The first device for

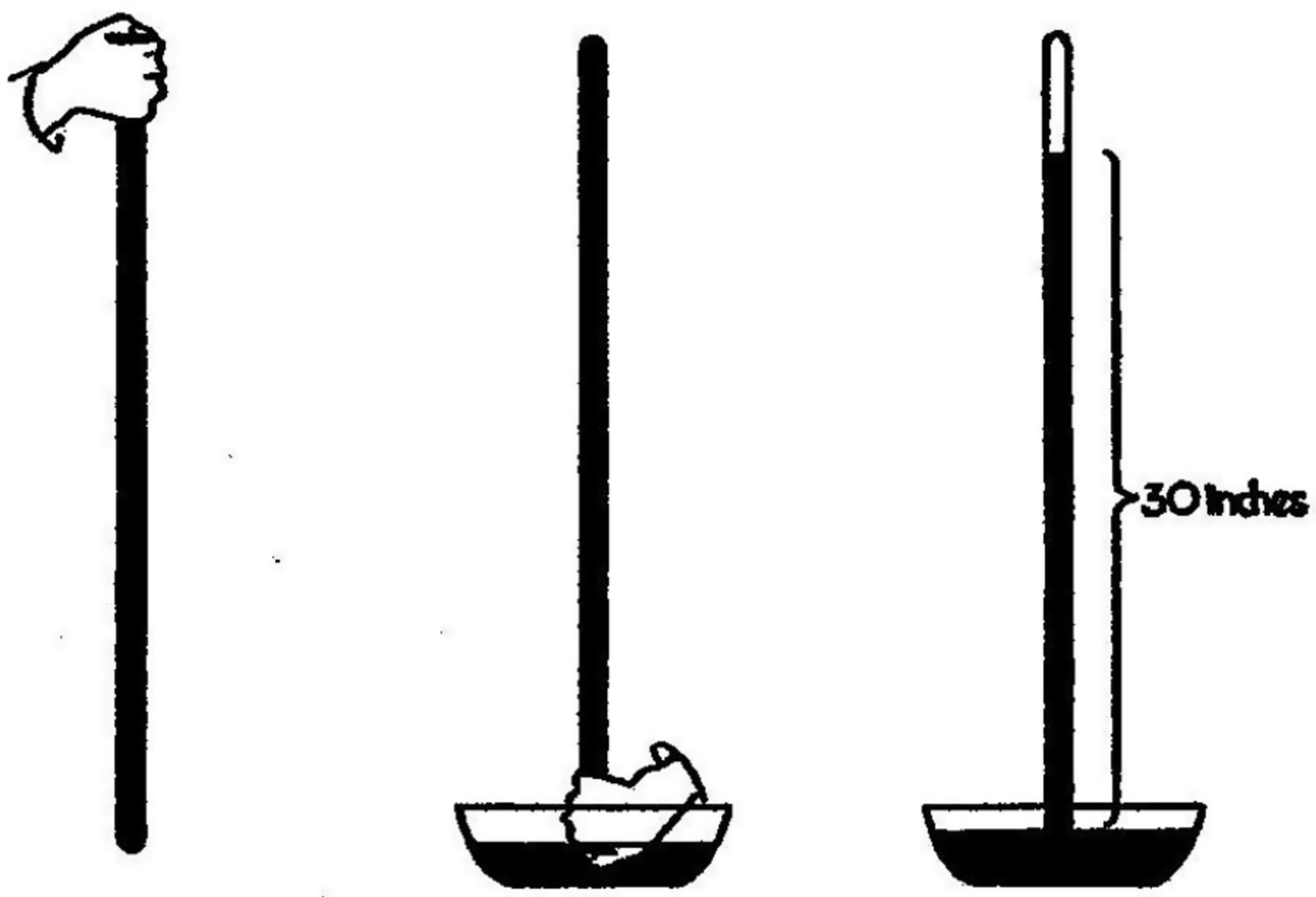


FIG. 83. Diagrams illustrating Torricelli's construction of the first barometer.

making such measurements was constructed in 1634 by an Italian scientist named Torricelli who had been one of Galileo's pupils. Torricelli's experiment is pictured in the three parts of Fig. 83. Taking a long glass tube which was closed at one end, he filled it with mercury. Then he inverted it in an open dish of mercury, while holding his thumb over the open end. When he removed his thumb he discovered that the mercury in the tube did not all run out but merely dropped to a point about 30 inches above the surface of that in the dish. In arriving at a tentative explanation of this phenomenon, Torricelli reasoned that since there was a vacuum above the mercury in the tube, and ordinary air above that in the dish, the 30-inch column of mercury probably just balanced, or weighed the same as, a column of air of the same cross-sectional area, reaching to the top of the atmosphere. Next he argued that if this were true and the device were

to be taken to the top of a mountain, the mercury column should fall since there would be less air above it to be balanced. Performance of this test experiment completely substantiated the theory. And so there came into being the first *barometer*. The mercury barometer of today as pictured in Fig. 84 operates on the same principle and has the same essential features as the simple one discussed above. More familiar to the average person is another type called the aneroid barometer. In this instrument, a flexible metal diaphragm which forms one side of a partially evacuated box moves back and forth with changes in the pressure of the atmosphere. These movements are transmitted through a system of levers to the pointer which registers on the circular dial. Even here the units on the dial are usually in terms of the height in inches of mercury. An aneroid barometer reading of 29 means that atmospheric pressure at that time would balance a column of mercury 29 inches in height. It is this type of instrument that is frequently given a different kind of dial and used as a " storm guide " or weather prediction device. The part that atmospheric pressure plays in weather forecasting was discussed to some extent in Chapter 7. One point may be emphasized again in this connection. While it is true that the kind of weather that we call bad is *usually* preceded by decreasing atmospheric pressure, yet it should be remembered that there are many factors which enter into the production of our weather. An instrument that reacts to only one of these cannot be expected to be completely reliable as a basis for weather forecasting.



FIG. 84. Modern mercury barometer.

It may also be noted that the common type of altitude measuring device or altimeter is simply a barometer, usually of the aneroid type. If, as already described, the barometer reading decreases as the instrument moves up a mountain, then the device could be used to measure the height of the mountain once the rate of decrease of the readings was known. The great disadvantage of such an altimeter when used in aviation is that it gives the height above the place from

which the plane started, and not that above the terrain over which it is flying at any given time. If the instrument was set at zero for the airport from which the plane started, a reading of 6,000 feet simply indicates an altitude of 6,000 feet above that airport. It is of no help in avoiding a 10,000-foot mountain that the plane may be approaching. There are, however, altimeters that register actual height above the portion of the earth over which the instrument is located.

Properties as a Liquid

Air in liquid form has come to be of considerable importance to man. Everyone is familiar with the fact that one way to change a gas to a liquid is to lower its temperature. When steam is cooled below 212° F., it changes to water. It will be recalled from an earlier discussion that on the basis of the kinetic theory, the difference between the gaseous and liquid states of a substance is primarily one of average distance between molecules, the separation being less in the case of the liquid. So it is obvious that increase of pressure on a gas should also help to transform it into a liquid, because this would force the molecules closer together. By combining low temperatures with high pressures all known gases have been liquefied. In the case of the mixture of gases which we call air the liquefaction point is about 320 degrees below zero Fahrenheit. Below this temperature it is a clear bluish liquid which has a density slightly less than water.

One of the most important uses of liquid air is in connection with the preparation in pure form of certain of its individual constituents. To illustrate the process, let us suppose that four liquids *A*, *B*, *C*, and *D* which have boiling points of 175° F., 125° F., 100° F., and 150° F. respectively, are mixed together and we wish to separate them. If the mixture should be heated to some temperature between 100° F. and 125° F., say 110° F., liquid *C* and *C* only will boil. If this vapor is collected and condensed, *C* will have been separated from the others. Then when all of liquid *C* has been evaporated, the temperature can be raised to a point between 125° F. and 150° F. and held constant while *B* vaporizes. Then, liquid *D* can be separated from *A*, completing the process. Since liquid air is just such a mixture of liquids with different boiling points its constituents can be separated in this manner. The process, which is called *fractional distillation*, is an

important commercial method of manufacturing several of the constituent gases of the atmosphere.

Liquid air also is used as a refrigerant in various laboratory procedures where very low temperatures are desired. In this connection it is worth noting that the properties of most substances change quite radically when they are subjected to such temperatures. Among the more spectacular demonstrations of such changes when the material is immersed in liquid air are the following. Mercury freezes and can be cast into a hammer, or other solid object. Alcohol also freezes and can be made into a candle which will burn much like the ordinary wax variety. A cigar or piece of charcoal dipped in liquid air and then lighted will burn explosively. Rubber becomes hard and very brittle. Lead wire becomes so elastic that it can be made into springs, while bells made from lead give off pleasing musical tones when at the temperature of liquid air.

Constituents

Oxygen. — In view of the fact that we have to breathe oxygen in order to live, in addition to being more than half composed of it, we can safely call it the most important constituent of air. Its discovery as a separate gas is usually credited to the English chemist-clergyman, Joseph Priestley, who first prepared it in 1774 by heating mercuric oxide. Priestley, incidentally, was an Englishman who strongly championed the cause of the American colonies and consequently was forced to flee to this country. In this way, the United States, because of the Revolutionary War, in a sense received one famous scientist in exchange for another, the Count Rumford who was mentioned in the preceding chapter.

Oxygen in its gaseous form is slightly heavier than air and will dissolve to a small extent in water. It is this dissolved oxygen that fish employ in respiration. As has already been mentioned, oxygen is used as the standard in our scale of atomic weights, being assigned a value of 16. It liquefies at a temperature of —297° F.

Oxygen occurs widely in nature making up some 21 per cent of the atmosphere, about 47 per cent of the earth's crust, and some 65 per cent of the human body. Except for the oxygen in the atmosphere, essentially all of it is chemically combined with other substances.

The general process which we call *oxidation* is the combining of oxygen with some material. When this reaction occurs so rapidly that light and comparatively large amounts of heat are given off, we give it the special name of combustion or burning. Cases of oxidation that proceed less rapidly include rusting or corrosion of metals, respirational processes, and organic decay. In all of them heat is being given off, the rate increasing with the speed of the reaction.

Oxygen may be prepared in the laboratory by heating mercuric oxide, the method used by Priestley, or by heating potassium chlorate ($KClO_3$) which has been mixed with manganese dioxide (MnO_2). The principal commercial method of preparing the gas is by fractional distillation of liquid air, as described earlier in this chapter. Some commercial oxygen is also made by a process known as electrolysis of water. In this method an electric current is passed through water as a result of which it breaks down into hydrogen and oxygen. The reaction is described chemically by the relation we used back in Chapter 13 as our first example of a chemical equation.

Most of the oxygen manufactured commercially is used with such devices as the oxy-acetylene torch in the welding and cutting of metals. It also finds certain applications in the manufacture of oils for paints, in various metallurgical processes, and to some extent as an explosive agent.

Nitrogen. — Nitrogen was discovered in connection with some experiments being made about 1772 on the portion of air that remained after animals had been confined in it for some time. Chemically, nitrogen is very inactive. It is a little lighter than oxygen, having an atomic weight of 14. In addition to making up almost four-fifths of the atmosphere, it is found in the earth in combination with certain other substances. The nitrate beds of Chile constitute an outstanding example of such a deposit that has great commercial importance. Nitrogen is prepared commercially by the fractional distillation of air and from deposits such as those mentioned above. Its principal use to man is as the inert element which dilutes the oxygen in his atmosphere. As a constituent of ammonia it is important in refrigerants and commercial fertilizers. Nitrogen also occurs in many kinds of explosives.

Carbon Dioxide. — This compound, which also was discovered by Joseph Priestley, composes about .03 per cent of the atmosphere.

Commercially it is prepared chiefly as a by-product of the fermentation industries. Important applications of carbon dioxide include its use in fire extinguishers, in carbonated water, as the medium by which bread and cakes are made "light," and as the modern refrigerant which we know popularly, though somewhat inaccurately, as "dry ice."

Rare Gases. — Included in this group are argon, helium, neon, radon, xenon, and krypton. Argon, which has an atomic weight of 40, is extremely inactive chemically and finds its principal use as a substance to put in electric light bulbs to reduce the rate of oxidation of the filament. It makes up about 1 per cent of the atmosphere. The others are present in air in very small amounts. The use of helium in balloons is familiar to everyone. A newer application for it is in connection with deep sea diving or other occupations where the worker is subjected to great pressures. It has been found that if such men are supplied while at work with a synthetic atmosphere of oxygen and helium instead of with ordinary air, the ill effects attendant on their return to ordinary pressures are greatly reduced. Neon finds its chief application in the neon signs of modern advertising. Radon is a gas that has radioactive properties like those of radium. No important commercial uses for xenon and krypton are known at the present time.

Water

The importance of water to man need not be further emphasized. In discussing it we shall proceed much as we did in the case of air, taking up first the material as a whole and then considering the one constituent of water that has not already been described.

Occurrence

Water may be said to occur essentially everywhere in nature. We have previously mentioned the fact that it covers something like three-fourths of the earth's surface and saturates a layer of rock under practically all the rest. It is sometimes spoken of as the Great Tourist because it falls to the earth as some form of precipitation, gathers into small streams which then flow into larger ones, eventually reaches lakes and oceans, and finally evaporates after which it is ready to start another round trip. The role played by water in the formation of

weather was discussed in Chapter 7. Water makes up an appreciable part of most materials and objects in the world, including ourselves. It constitutes, for instance, some 70 per cent of the human body, about 90 per cent of the brain, approximately 60 per cent of beef, 85 per cent of milk, 90 per cent of watermelon, and 95 per cent of cucumbers.

Properties

Water is a compound, the chemical formula for which, H_2O, we have already used. Being a compound it has the property of definite composition. In pure form at ordinary temperatures it is a clear, odorless, tasteless liquid. Its density is such that 1 cubic foot of it weighs about 62.4 pounds, or 1 cubic centimeter, 1 gram. The latter figure, as was pointed out previously, was originally used in defining the gram as a unit of mass. On the Fahrenheit temperature scale water freezes or solidifies at 32° F. and boils or vaporizes at 212° F. It will be recalled that its freezing and boiling points were used as the defining temperatures of the Centigrade scale, being called 0° C. and 100° C. respectively. In its pure form it is a very poor conductor of both heat and electricity. Lest one infer from this statement that there is no truth in reports that people have been severely shocked or electrocuted as a result of touching a defective light switch while standing in the bathtub, it should be emphasized that the above applies to water which contains no impurities. Various materials dissolved in water render it electrically conducting and the danger increases with the area of the body that is made conducting.

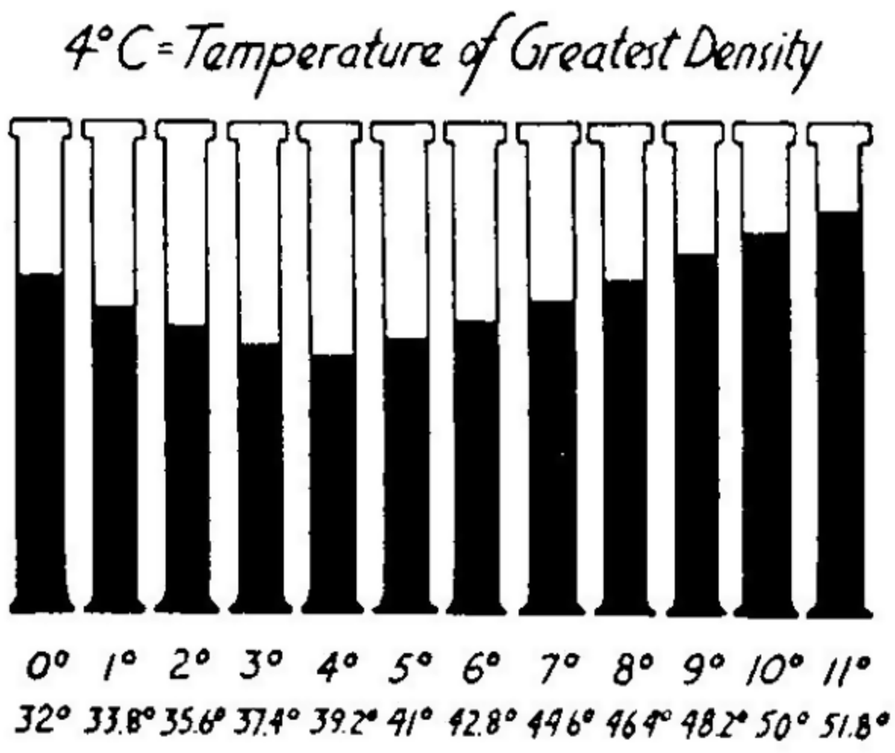


FIG. 85. Diagram showing schematically the manner in which the volume of a given mass of water varies with the temperature.

The experimental fact that, in general, substances expand with rising temperature was discussed earlier in this book. Water is of particular interest in this connection because of its rather peculiar behavior. This behavior is illustrated in Fig. 85. At the left of the diagram there is shown a glass graduate containing water which is just above freezing. From

left to right in the figure there is pictured the volume which this water occupies at increasing temperatures. It will be noted that as the temperature rises from freezing to 4° C. or 39.2° F., the water contracts. At 4° C. it occupies the least volume, expanding again from 4° C. on up. Thus 4° C. or 39.2° F. may be said to be the temperature of greatest density for water. In other words, it is the temperature at which any given volume of water has the greatest weight. In the diagram in Fig. 85 the amount of the volume change is somewhat exaggerated in order to make the phenomenon clear. Actual data were used in plotting the graph shown in Fig. 86. Here the same sort of result is indicated in more conventional form.

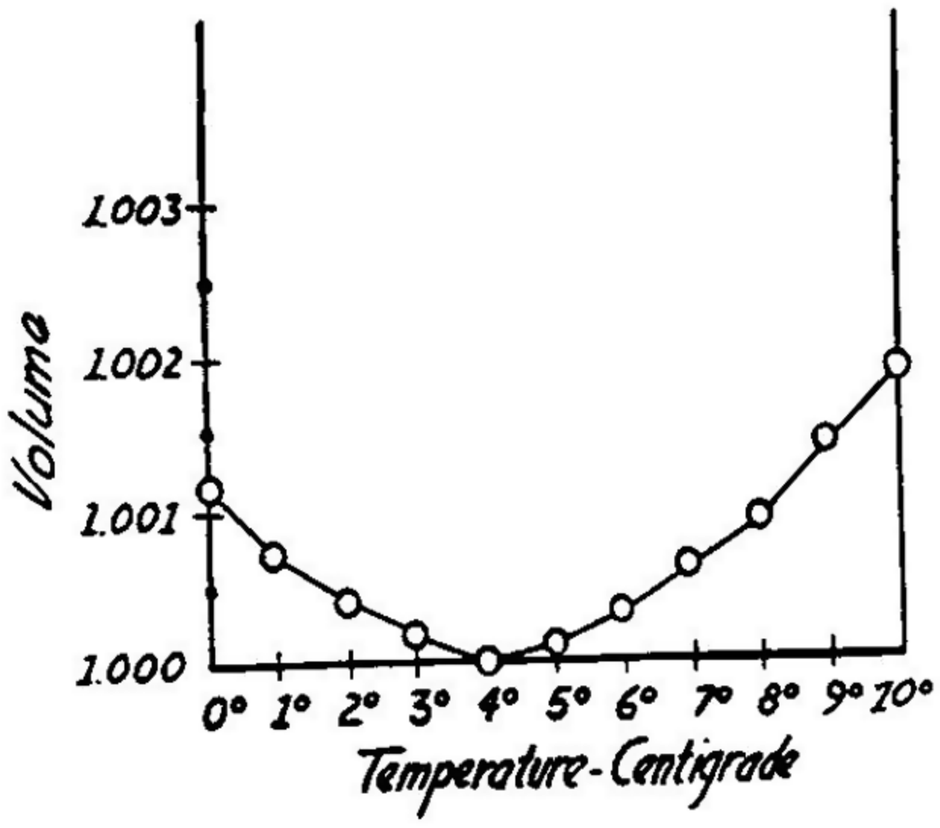


FIG. 86. The conventional volume-temperature curve for a given mass of water.

Another unusual characteristic possessed by water is its behavior during solidification. It was pointed out in Chapter 8 in connection with the discussion of the erosional work of water that when it freezes it expands to approximately 11/10 of its liquid volume. Most common substances, on the other hand, contract during solidification instead of expanding. One of the reasons that gold and silver coins are stamped out instead of being made in molds is the fact that as these metals change from liquids to solids, they contract. This means that they would draw away from the corners of the mold and fail to reproduce accurately the desired design. The rather peculiar behavior of water in this respect has several important consequences for us. Among the more undesirable ones are the results of its freezing in water pipes, china pitchers, or automobile radiators. Fortunately, these are more than outweighed by one other result. If water contracted upon freezing, ice would have a greater density than water and so would sink to the bottom of a lake or pond instead of floating on the surface. Such bodies of water would freeze from the bottom up instead of from the top down. Since water is a poor conductor of heat the sun's rays would fail to heat these bodies to any great depth, and most of them would have a layer of ice at the bottom throughout

the entire year. Such a state of affairs would have very serious effects on the existence of life upon the earth.

One other property of water that should be mentioned is the fact that it is the best solvent known. In other words, larger amounts of more different things will dissolve in water than in any other known liquid. In the annals of alchemy there is an ancient story to the effect that many men spent long years looking for a universal solvent, that is, a liquid that would dissolve indefinitely large amounts of all materials. According to the tale, the search came to an end only when some individual became curious as to what the alchemists expected to use as containers for this wonderful fluid after they found it. No such magic liquid has ever been discovered, but water comes nearer than any other to meeting the qualifications.

Chemically water is a very stable compound. It does not readily break down into its constituents even under extreme conditions. Even when raised to a temperature of 3,600° F., less than 2 per cent of water molecules are dissociated into their constituent parts. Water combines with certain oxides of non-metals, such as carbon dioxide (CO_2), and sulfur dioxide (SO_2) to form acids. In the first of these, the product is carbonic acid (H_2CO_3), one of the erosional agents discussed in Chapter 8.

Composition

The ancient world thought of water as an element along with fire, air, and earth. It was not until some years after Priestley's discovery of oxygen, that the true nature of water became known. About 1795 Henry Cavendish performed an experiment in which an electric spark was generated in a closed vessel containing a mixture of hydrogen and oxygen. He found that water was formed and that the gases had been used up in the ratio of two volumes of hydrogen to one of oxygen. Thus, water was established as the compound which we designate by the formula H_2O. The only other common compound of hydrogen and oxygen is hydrogen peroxide, which has the formula H_2O_2. At one time this substance was thought to be an excellent antiseptic, but it was probably considerably overrated in view of the fact that it is very unstable and rapidly breaks down into water and oxygen.

Many materials which are not part of the compound are found

in natural waters. The relative total amounts of these substances as they occur in several different kinds of natural water are shown in Fig. 87. It will be noted that even rain water, which one ordinarily thinks of as being comparatively pure, contains appreciable amounts of foreign materials. These include dust and bacteria in suspension, and carbon dioxide (CO_2), sulfur dioxide (SO_2), ammonia (NH_3), nitrogen, and oxygen, in solution. The materials found in spring and well water depend on the composition of the rock layers which the water has encountered in flowing over and through the earth's crust. Consider, for example, some rain water which has acquired a certain amount of carbon dioxide in solution. This constitutes carbonic acid, which will attack limestone. Therefore if the water flows through limestone, a reaction occurs which results in the formation of calcium bicarbonate, the mineral found most frequently in so-called "hard" water. Other materials commonly present in spring and well water include the carbonates and sulfates of magnesium, sodium, potassium, iron, and aluminum. Oxides of silicon, iron, and aluminum often occur in suspension. River water generally contains, in addition to the above, organic matter which results from decay. Ocean water, as the diagram indicates, contains on an average something like eighty times as much dissolved matter as does deep well water. It has been estimated that every cubic mile of ocean water contains something like 175,000,000 tons of chemical combinations of gold, silver, copper, magnesium, aluminum, calcium, bromine, iodine, and other materials, having a total value of around five billion dollars. Because of the extremely low concentration, with one exception, it has not been economically feasible thus far to try to extract these substances. The exception is bromine, which for some years now the Dow Chemical Company has been removing from ocean water in a plant located in Wilmington, North Carolina. As a by-product of bromine production enough gold has been taken from the waters of the sea to make one small pellet. The company's success with bromine extrac-

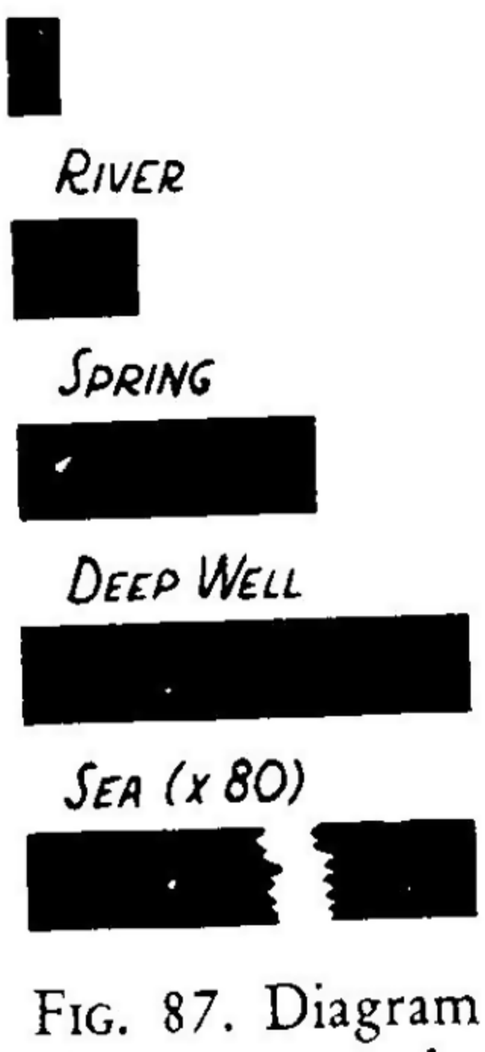


FIG. 87. Diagram illustrating the comparative amounts of substances dissolved in various kinds of natural water.

tion has led it to plan the construction of a plant to remove magnesium from the waters of the Gulf of Mexico.

Uses by Man

As a Solid. — Water is one of the few substances that man knows familiarly and makes use of in all of its three states. As the solid that we call ice, it finds its principal use in the field of refrigeration. Francis Bacon is credited with having been among the first to suspect the possibilities of ice in the preservation of food. There is a story to the effect that one day in 1626 while on a trip he had the coachman stop at a farmhouse where he purchased a chicken, killed and dressed it, and then stuffed it with snow to discover whether in this way it could be preserved for a longer time than usual. From the standpoint of refrigeration of food the experiment was a complete success. Bacon, however, as a result of the investigation, is believed to have contracted bronchitis from which he died. The ice industry in the United States attained no importance until a little after 1800. At that time a yellow fever epidemic was raging in the West Indies. It occurred to a wealthy New Englander, named Francis Tudor, that a boatload of ice would be very useful to the doctors engaged in battling the disease. In spite of the ridicule of his friends, all of whom were quite certain that the ice would be entirely melted long before it reached the West Indies Tudor proceeded to load a boat with ice and start it off. He had made something of a study of methods of preserving ice by proper insulation and the venture was very successful, both as an aid in combating the epidemic, and as the beginning of a profitable tropical ice business.

As a Liquid. — Man's uses of water, the liquid, are primarily three in number, namely, for drinking, washing, and the production of steam in boilers. The impurities we have already mentioned, in addition to various others, may render natural water unfit for any or all of these purposes, and so must be removed.

In the case of drinking water the chief problem is to guard against disease-producing bacteria. Since these bacteria can be seen only with the aid of a microscope, a clear and sparkling appearance cannot be taken as a guarantee of purity in water that is to be used for drinking purposes. Most water-borne diseases are caused by bac-

teria which get into the water with sewage. For this reason the importance of water purification increases as centers of population become more and more crowded. Water that is absolutely pure can be prepared only by distillation. However, since bacteria causative of disease are not dissolved in the water but simply carried in suspension, cheaper methods of purification can be employed. The two processes which are used on a large scale are filtration and chlorination. The filters are usually large beds of sand or gravel which strain out the bacteria as the water seeps through. Certain chemicals are often added in order to speed up the process. Treatment of drinking water with chlorine constitutes the cheapest method of purification where large quantities of water are desired. Although the process was not used in this country until 1913, today more than half of the country's population drink chlorinated water. The chlorine, which acts as a poison to the bacterial cells, need not be present in concentrations greater than 7 parts per 10 million gallons of water. It is largely because of the adoption of purification practices that typhoid fever has been reduced from a major to a minor disease in America. Deaths from it in the United States have declined from about 36 per 100,000 population in 1900 to 2.8 in 1935, and in Canada from 10 in 1921 to 2.3 in 1936. Illness and death from other water-borne diseases have decreased similarly.

The impurities which render water unfit for washing and steam production are dissolved materials, principally salts of calcium and magnesium. Water which contains these substances in comparatively high percentages is said to be "hard" and the process of softening consists in removing them. If water can be softened by the application of heat alone it is said to be of temporary hardness. Such waters contain the magnesium and calcium salts known as bicarbonates. Heating the water causes decomposition as illustrated in the following equation for calcium bicarbonate.

$$\underset{Ca(HCO_3)_2}{\text{Calcium bicarbonate}} \rightarrow \underset{CaCO_3}{\text{Calcium carbonate}} + \underset{H_2O}{\text{Water}} + \underset{CO_2}{\text{Carbon dioxide}}$$

The calcium carbonate is insoluble in water and so sinks to the bottom of the container from which it can be removed. The sulfates of calcium and magnesium cannot be removed from water by heating

and so such water is said to be of permanent hardness. It can be softened by chemical treatment only. The disadvantage of dissolved materials from the standpoint of washing is simply that a lather cannot be worked up in water containing them. When such water is turned into steam in boilers, the compounds that were in solution are deposited on the inside of the boiler tubes tending to clog them up and reduce the boiler's efficiency. It has been calculated that as little as a quarter of an inch of such scale will increase fuel consumption by something like 50 per cent.

In addition to its uses in washing, drinking, and steam production, water serves man in at least two other ways that should be mentioned. One of these is the very practical service it gives as the substance upon which he floats the boats of the world. The other is as a sort of scientific yardstick. We have seen already that the Centigrade scale of temperature uses the freezing and boiling points of water as its fixed temperatures, that definitions of the gram and ounce are in terms of water, and that units of heat, such as the calorie and British thermal unit, are also so defined. In addition to these, water is used as the standard substance in defining specific heat and specific gravity, which are respectively, measures of relative rates of heating and relative densities of substances.

As a Gas. — Water in its gaseous form, steam, serves man chiefly in two ways. The more important of these is as the working substance in the steam engine, one of the devices by which man transforms heat energy into mechanical energy. This application will be considered in some detail in a later chapter on heat engines. The other use is in the heating of buildings. As we saw in Chapter 13 considerable quantities of heat are required to change water into steam and these same amounts are released when the steam condenses. In this way a building may be heated by converting water into steam in the basement and then conducting that steam through pipes to radiators located in the various rooms, where the condensation occurs.

Heavy Water

A substance called "heavy water" has received a considerable amount of scientific attention and some popular publicity in recent years. Our discussion of it at this point necessarily will be rather in-

complete, since the manner in which it differs from ordinary water is one of internal atomic structure. It will be recalled that hydrogen was said to have an atomic weight of 1.008 and the impression was probably left with the reader that all hydrogen atoms are alike. For most purposes such an idea is quite satisfactory, but here is a place at which we must be a little more exact. As a matter of fact, there are three separate and distinct kinds of hydrogen atoms having atomic weights of 1, 2, and 3 respectively. Since considerably more than 99 per cent of the total consists of the first kind, the average atomic weight of all is just a little more than 1, namely the 1.008 mentioned above. To be exact, about one in 5,000 has a weight of 2. Although the " atomic weight 3 " variety does not exist naturally, it can be made artificially. The reader with some knowledge of chemistry will recognize that here we are describing isotopes of hydrogen. The question as to why atoms that weigh almost two and three times as much as the average should still be called hydrogen is one that will have to be saved for a later chapter in which various matters concerning atomic anatomy are discussed. It will be obvious, however, that if a molecule is composed of one atom of oxygen and two atoms of hydrogen of the " atomic weight 2 " variety, it will be somewhat different from a molecule of ordinary water. This " heavy hydrogen " is called deuterium and the heavy water formed by its combination with oxygen designated by D_2O to distinguish it from H_2O. The density of heavy water is about 10 per cent greater than that of ordinary water. It freezes at 3.8° C. and boils at 101.4° C. It has been shown to be poisonous to many plants and animals. Various seeds will not sprout in it. Fish, tadpoles, and worms have been known to die from drinking it. Several investigators have tested its effects on themselves by drinking small amounts of it daily. The only noticeable result has been a slight slowing down of the metabolism. This has led to the semi-serious suggestion that individuals might possibly prolong their lives a little by drinking heavy water as they grow older.

Hydrogen

Of the two constituents of water, one was discussed earlier in this chapter. The other, hydrogen, was discovered in 1766 by the English scientist, Henry Cavendish. He found it could be prepared by pouring either sulfuric or hydrochloric acid on iron, tin, or zinc.

Hydrogen is the lightest of all the elements, having an atomic weight of 1.008. It is highly inflammable. It liquefies at —422.8° F. and solidifies at —434.45° F. Hydrogen occurs very widely in nature, mostly in chemical combination with other substances. In addition to being present in water, it is a constituent of all carbohydrates, fats, proteins, alcohol, acids, and many other compounds. It will be recalled from the discussion in Chapter 7 that at high altitudes free hydrogen makes up a high percentage of the atmosphere.

Most of the hydrogen used industrially is prepared either by electrolysis of salt water, or by passing steam over heated carbon, usually in the form of coal. Commercial uses of the gas are varied in nature. One is in the process called hydrogenation. Applied to the manufacture of vegetable cooking fats, this consists in bubbling hydrogen through a liquid fat such as cotton-seed oil. The resulting reaction forms a semi-solid fat of the sort that is so widely used today in cooking in place of lard. Another application of hydrogenation is in the manufacture of synthetic gasoline, a process that is discussed in the next chapter. Hydrogen also finds some use as a so-called reducing agent. Chemically, the process of *reduction* may be defined for our purposes as essentially the opposite of oxidation. That is, reduction consists in removing oxygen from a substance. If hydrogen is passed over hot copper oxide (CuO), the reaction described below occurs,

$$\text{Copper oxide} + \text{Hydrogen} \rightarrow \text{Copper} + \text{Water}$$
$$\text{CuO} + \text{H}_2 \rightarrow \text{Cu} + \text{H}_2\text{O}$$

It will be noted that the hydrogen acts as an agent which removes the oxygen from the copper oxide, thus reducing it; also, the hydrogen was oxidized. It is true in general that any reaction which involves one of these processes also involves the other. In Chapter 17 on Metals and Alloys we shall find that reduction is a very important metallurgical process.

Questions

Discussion

1. The ancients considered air to be one of four elements; what were the other three?

2. Describe in detail the experiment by which Torricelli demonstrated that the atmosphere exerts a pressure.

3. Explain just how Torricelli could have used his mercury barometer as an altitude-measuring device.

4. Suppose an aneroid barometer which reads 30 is placed in a jar that can be evacuated; what changes would you expect to see take place in the instrument's reading as the air is pumped out of the container? Explain.

5. The reading of 30 referred to in the preceding question would be in what kind of units?

6. Discuss the barometer as an instrument on which to base weather predictions.

7. Explain fully the process of fractional distillation.

8. Look up some additional facts about the life and work of Joseph Priestley.

9. List the various elements in the atmosphere in order of increasing atomic weight.

10. Define oxidation and relate it to combustion and corrosion.

11. Find out what the phlogiston theory was and why it was abandoned.

12. Look up information about one or two commercial uses of oxygen other than those suggested in this chapter.

13. Find out what sublimation is.

14. Find out something about the manufacture of solid CO_2 or " dry ice," preferably by visiting a plant where it is made.

15. Reconcile the statement, " CO_2 is the medium by means of which bread is made 'light,'" with the one, " Part of the efficiency of CO_2 as a fire extinguisher lies in the fact that it is heavier than air."

16. At what temperature on the Fahrenheit scale would a quart of water weigh the most? Explain.

17. Does the fact that water is used in hot-water heating systems imply that it must be a good conductor of heat? Explain.

18. Distinguish between so-called " hard " and " soft " water.

19. Discuss two methods of large-scale water purification.

20. Name as many scientific quantities as you can whose definitions depend in some manner or other on water.

21. Find out something about the water system of some large city.

22. Describe in detail the water system of your own locality.

23. How can a gas as inflammable as hydrogen exist in the atmosphere of a body which is at as high a temperature as is the sun?

24. Define reduction and find one use of hydrogen as a reducing agent other than the one mentioned in this chapter.

25. Find one example of hydrogenation other than the one described in this chapter.

Multiple Choice

1. The reading of a mercury barometer entirely enclosed in a vacuum will be — (a) 100, (b) 30, (c) 20, (d) 15, (e) 0 — inches of mercury.

2. To which one of the following might one prefix the adjective " slow "

nd have a synonym for "rusting"? (a) Reduction, (b) Electrolysis, (c) Oxidation, (d) Phlogistonation, (e) Calcification.

3. At which one of the following temperatures would 50 pounds of water ccupy the least space? (a) 1° C., (b) 50° F., (c) 50° C., (d) 40° F., e) 99° C.

4. List all of the following that are true of the equation:

$$CuO + H_2 \rightarrow Cu + H_2O$$

a) Represents an important laboratory method of preparing hydrogen. b) Represents an important commercial method of preparing hydrogen. c) Is balanced. (d) Involves both reduction and oxidation. (e) Involves either reduction nor oxidation.

5. Water (one or more) (a) may be said to be a mixture of hydrogen nd oxygen, (b) possesses the property of definite composition, (c) in a pure orm is a poor conductor of heat, (d) contracts as it cools from 95° C. to ° C., (e) is the only known substance composed of just hydrogen and xygen.

True–False

1. All oxidation processes generate heat.

2. A mercury column approximately 60 inches in height would be necesary in order to balance a pressure of 22 pounds per square inch.

3. Liquid air boils at absolute zero.

4. All combustion is oxidation, but not all oxidation is combustion.

5. Argon, neon, krypton, xenon and helium together constitute less than per cent of the atmosphere.

6. More than 50 per cent of the reader is water.

7. Water of permanent hardness is differentiated from that of temporary ardness on the basis of the number of bacteria present per unit volume.

8. The principal reason for the use of chlorination in addition to filtraion is that the latter will not remove disease-causing bacteria which are disolved in the water.

9. Reduction and hydrogenation are synonymous terms.

10. A molecule of heavy water contains four atoms instead of three as in he case of ordinary water.

Chapter 16

FUELS

THE period in which we live has been given many different titles by people who wished to emphasize one phase or another of modern life. Such names as Electrical Age, Iron Age, Age of Rubber, Age of Plastics, are heard frequently. A good case can be made for Age of Fuels as being as descriptive of the times as any other. Everyone recognizes today that fuels can no longer be thought of solely, or even chiefly, in terms of cooking food and heating buildings. Ever since the first wood-burning steam engine was adapted for use in a boat or a wheeled vehicle, transportation has been growing increasingly dependent on fuels. Today most of our methods of moving either materials or ourselves from one place to another make use of gasoline, oil, or coal. Also, a large amount of the electrical energy used today can be traced back to the engine that operates the generator and thence to some fuel, the combustion of which enables the engine to deliver energy. In fact, any operation in which the energy employed comes from a gas engine, a Diesel engine, a steam engine, or similar device, is dependent on a fuel. Life itself could not go on without fuel, because food can be so classified. The processes which transform foodstuffs into energy and body building materials are essentially the same as those that occur in ordinary fuel combustion.

In this chapter it is our purpose, first to expand the general picture of combustion a little further than was possible in Chapter 15, and second to discuss a number of specific materials that we use as fuels. With regard to the latter we will consider origin, heating value, by-products of manufacturing processes, uses other than as fuels, and resources.

Factors Influencing Combustion

Review of Definition

It will be recalled from the earlier discussion of combustion that when a substance is burning, oxygen is uniting with it at a rate rapid enough to produce light and a reasonably large amount of heat. In other words, combustion is pictured as high speed oxidation. With this fundamental picture in mind, we will consider certain factors that exert an influence on the rate at which the process of combustion takes place.

Effect of Temperature

Two experimental facts are chiefly responsible for the effect that temperature has on a burning fuel. In the first place, the rate of burning or combustion increases as the temperature of the burning substance rises, a relationship that holds for most chemical reactions. In the second place, oxidation liberates heat, the rate of liberation increasing with the temperature. Now it will be observed that the two facts just mentioned lead to a kind of vicious circle which may be stated as follows: The faster a substance burns, the more heat it gives off; the more heat it gives off, the higher its temperature becomes; the higher its temperature becomes, the faster it burns; the faster it burns, the more heat it gives off; and so forth.

Spontaneous combustion is an outstanding example of the sort of effect that this accumulative process can have. Any combustible material, under a given set of conditions, has what may be called its kindling or ignition temperature. That is, it must be heated to a certain temperature before it will burn. A newspaper lying on the living room floor is not likely to suddenly burst into flame and burn down the house. This is not because there is no oxidation of the paper; it is simply due to the fact that the heat liberated by the oxidation is conducted or convected away as fast as it is generated. The newspaper is not heated to its ignition temperature. But if some rags soaked with oil or paint are thrown in a heap under the cellar stairway or in a corner of a clothes closet, the ventilation may be so poor that the heat liberated by oxidation is not carried away as rapidly as it is formed. The heat of oxidation raises the temperature of the rags, which in turn

increases the rate of oxidation, which raises the temperature still further, and so on, as indicated above. Eventually they will be very likely to reach kindling temperature and begin to burn, and the following morning the newspapers will carry the story of another home destroyed by fire, believed to have started by spontaneous combustion. Such fires also often begin in freshly cut hay which is put in the barn while still damp, or in coal that is stored in poorly ventilated coal bins or even outdoors.

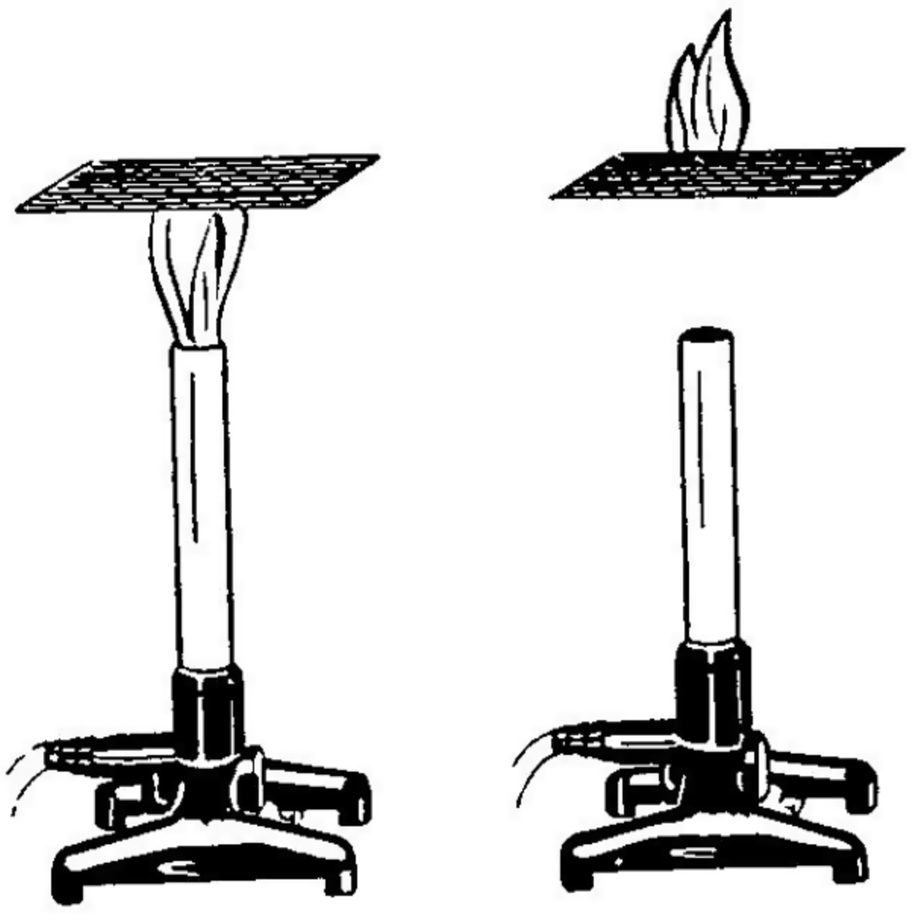


FIG. 88. Diagram to illustrate the fact that a combustible substance must be raised to its kindling temperature before it will take fire.

The fact that a combustible material does not burn until it reaches its kindling temperature may be demonstrated in the experiment pictured in Fig. 88. Gas from a Bunsen burner is allowed to rise through a wire mesh. As shown, the gas above the mesh will burn without that below being ignited, or that below can be lighted and the portion above not take fire. This is simply because the wire conducts the heat away so rapidly that gas on the other side of the mesh is not raised to its ignition temperature. A miner's safety lamp developed by Sir Humphry Davy makes use of this phenomenon. In it, the flame is surrounded by a metal gauze which permits light to shine out and oxygen to enter to maintain the combustion, but prevents the air outside from reaching a high enough temperature to ignite any combustible gases that may be in it.

Effect of Surface

In the case of a solid fuel, the oxidation reaction occurs at the surface which is the boundary between the air and the combustible material. Other things being equal, the rate of combustion is proportional to the area of this interface. Breaking a large piece of a solid fuel like coal into smaller pieces increases the ease with which it can be ignited and the speed with which it burns, since by this pulverizing process the surface area is greatly increased.

That this is true is evidenced by various everyday experiences, such as the fact that shavings burn more readily than do logs, or that loose papers can be ignited more easily than can tightly baled papers. Somewhat more startling evidence is available in connection with materials which we do not ordinarily think of as being combustible. Iron bridges and aluminum kettles are not generally considered to be highly inflammable. If, however, the iron or aluminum is powdered and then blown into a gas flame, it burns very readily. Powdered coal burns explosively if the particles are kept from packing together and thus reducing the surface in contact with the air. As a matter of fact, the first Diesel engine was designed to run on powdered coal rather than on oil, and was so successful from the standpoint of the generation of power, that it blew up.

It is this great increase in surface which results from reducing particle size that sets the stage for dust explosions in industrial plants. The fact that such explosions have occurred in dust of leather, spice, starch, grain, cocoa, sugar, cork, wood, aluminum, bronze, coal, dry milk, and other materials, indicates that the danger is not confined to factories dealing only with substances which we commonly picture as highly inflammable.[1] The evidence indicates that such explosions do not occur spontaneously but must be set off by some means of ignition such as a flame or a spark. The danger of explosion is increased by the fact that the ignition point of some dust is not extremely high, being, in fact, well below the temperature that would cause iron to glow a dull red.

Effect of Concentration

The factor of concentration may be considered from two standpoints, namely, that of the oxygen in the atmosphere, and that of the fuel itself. With respect to the former, it is true that the rate of oxidation in general increases as the percentage concentration of oxygen becomes higher. It is because of this that an oxyacetylene welding torch, in which acetylene gas is burned in an atmosphere of high oxygen content, generates more heat than does acetylene gas when it burns in ordinary air.

[1] Brown, H. R., "Dust Explosions in Industrial Plants," *Ind. & Eng. Chem.*, 17, 902 (1925).

It will be obvious that the concentration of the fuel itself, in the mixture of fuel and air, can be varied to any great extent only in gaseous fuels. For most efficient combustion there must be neither too low nor too high a concentration of gas in the mixture of gas and air. Too little gas is unsatisfactory for obvious reasons. On the other hand, if the percentage of gas is very high, there will not be enough atmospheric oxygen present to produce complete combustion. The best mixture of the two substances varies with the type of gas. For natural gas the most efficient concentration of the fuel is from 6 per cent to 13 per cent.

Solid Fuels

Coal

Coal as we find it today may be said to be the carbonized remains of what was luxuriant vegetation hundreds and thousands of centuries ago. Coal mined in western Pennsylvania experienced its earlier existence in the form of trees, huge ferns, or club mosses, during the latter part of the Paleozoic Era, some 200,000,000 years ago. Although both wood and coal contain carbon, hydrogen, and oxygen, they differ in that the percentage of carbon in the latter is much higher than in the former. This has come about, not by the addition of carbon during the process of coal formation, but by the forcing out of oxygen and hydrogen. The family history of an average load of coal is somewhat as follows: The vegetation that was ancestral to it grew in great profusion in swampy regions. As trees died and fell over or were blown down, they were covered with water and mud, which retarded the normal processes of decay. As time went on the weight of material above them became greater and greater as a result of which they were subjected to large pressures and often to high temperatures. The effect of the high pressures and temperatures was to force out of the wood a number of gases containing hydrogen and oxygen, leaving the remains more and more nearly pure carbon. Among these gases are carbon monoxide (CO) and methane or marsh gas (CH_4). The particular kind of coal that our average load happens to be, depends on how far the process described above had progressed when the miner interrupted it by digging up the coal. The percentage composition of wood and of the various kinds of coal together with

their average heat values in British thermal units per pound are given in Table 14.

TABLE 14. COMPOSITION AND HEAT VALUE OF VARIOUS FUELS

Fuel	*Carbon*	*Hydrogen*	*Oxygen*	*Heat Value*
Wood	51%	6%	43%	8,000–9,000 B.t.u./lb.
Peat	58	6	36	6,500–9,500
Lignite	70	5	25	9,000–14,000
Bituminous	88	5	7	14,500–15,000
Anthracite	94	3	3	14,500–16,000

Peat, the first stage in the process, is formed before pressures of very great proportions have been exerted. It ignites easily and burns with a hot flame. Peat finds its principal use as a fuel in countries where better grades of coal are not available. Although there are large deposits of it in Minnesota, Virginia, North Carolina, and Georgia, almost none is burned in this country.

After the carbonizing process has continued somewhat longer the coal is called *lignite*. This material is harder than peat, more coal-like in surface appearance and usually dark brown in color. Like peat it finds little use in this country because of the abundance of better fuels. There are, however, large amounts of lignite available to us. Layers varying from 4 feet to 35 feet in thickness cover some 35,000 square miles of South Dakota. The use of lignite by various European countries in the manufacture of synthetic gasoline will be discussed later in this chapter.

After lignite comes *bituminous* or soft coal, of which enormous quantities are used both in this country and in Europe. It has been estimated that as much bituminous coal has been burned in the last ten years as in all previous history. The United States is particularly blessed with respect to its resources of this particular fuel, having something like six times as much as South America, three times as much as Europe, and between two and three times as much as Asia. Much of our use of bituminous coal is extremely wasteful in that many valuable materials contained in it are completely lost, a point that will be covered in some detail in connection with the consideration of coke.

Anthracite, or hard coal, which is from 90 per cent to 95 per cent carbon was not discovered until 1791, and did not come into important industrial use until after 1820. The difficulties experienced in

getting it to burn and the fact that boilers designed for use with wood or soft coal fires were unable to stand up under the high temperatures generated by hard coal, accounted for this delay. To complete the story of nature's carbonizing process, it should be mentioned that it does not stop with anthracite coal. At the end of the line is *graphite*, a material that is almost pure carbon. It is not used as a fuel but finds important applications as a lubricant and as the substance from which " lead " pencils are made.

Coke

Coke is a manufactured product which is made by heating bituminous coal in a closed container, that is, in the absence of air. The process is called *destructive distillation.* When the coal is treated in this fashion it begins to grow soft and spongy at temperatures between 600° F. and 700° F. At still higher temperatures a number of gases and vapors are given off, until at the end of the distillation process only a grayish-black porous substance is left in the oven. This material, which is almost pure carbon and has a heat value in the neighborhood of 14,000 B.t.u. per pound, is coke. Its chief use has been in the past, and still is, as a reducing agent in the metal industries, particularly in the manufacture of iron. Something like one seventh of all the coal mined in the United States is used in this manner. Today, however, increasing amounts of coke are being employed in place of coal in heating homes and other buildings. Until about 1910 essentially all the coke manufactured in the United States was made in a device known as a beehive oven. Since the manufacturers were interested in nothing but the coke, no attempt was made to save any of the vapors which were given off in the course of the coking process. The oven was simply filled with soft coal, the lower layers of which were burned. Thus heat was provided for transforming the upper layers into coke. The volatile gases escaped through the opening in the top of the oven. The fact that valuable substances were being lost in these gases was first recognized in Europe a little after 1900, and attempts were made to design coke ovens which would save them. These by-product ovens were not introduced into the United States until 1910 and did not produce even as much as half of the country's output of coke until almost 1920. Today more than

90 per cent of our coke is manufactured in ovens which save practically every constituent of the coal that is given off during the distillation process.

The by-product coke oven proper is a long narrow rectangular chamber some 30 feet long, from 6 to 10 feet high, and a foot or two wide. Each oven is located between two heating chambers which serve to heat the soft coal with which the ovens are filled. The fuel burned in the heating chambers is a mixture of gas and air which has been pre-heated in the lower chambers. Oven temperatures are in the neighborhood of 1,800° F. Volatile gases driven from the coal are conducted off through a flue in the top of the oven. When the coking process is completed, the ends of the oven are opened and the coke pushed out into cars by means of a huge ramrod. In some cases batteries contain as many as 600 of these ovens, each of which may be able to handle some 8 tons of coal every 20 to 24 hours.

Now let us see just what soft coal can yield in the way of by-products when it is properly treated. It has been found possible to get approximately the following from one ton of Pittsburgh coal: [2]

1,500	pounds of coke
11,360	cubic feet of coal gas
12.5	gallons of coal tar
25	pounds of ammonium sulfate
4	gallons of light oil

The uses of coke, as has already been pointed out, are as a reducing agent in the metal industries, and as a fuel. The coal gas, which will be considered again later in the chapter, is a valuable fuel containing some 97 per cent of combustible materials.

The list of modern materials whose source is coal tar is an impressive one, and includes dyes more numerous and of better quality than can be obtained from natural sources, numerous drugs, saccharine, explosives like T N T and its relatives, certain insecticides and wood preservatives, synthetic plastics, flavoring agents, carbolic acid, photographic developers, and solvents. Ammonium sulfate provides a source of certain fertilizers, calico dyes, and other products, as well as ammonia itself.

[2] "Chemistry in Industry," Vol. I, Chap. IV; Chemical Foundation.

The above facts make it obvious that our common practice of burning soft coal in the furnace is an extremely wasteful procedure. This has led to the suggestion that the ideal way to use bituminous coal would be to " coke " it at the mine. The coal gas then could be transported by pipeline and used for heating and cooking. The coke also would be available for use as either a fuel or reducing agent, and all the other valuable by-products would be saved.

Charcoal

When the process of destructive distillation, described above, is applied to wood, the solid product is called wood charcoal. It is almost pure carbon and results from water being driven off from the cellulose.

Materials other than cellulose in the wood break down into a series of volatile gases from which valuable by-products are obtained. Other forms of this fuel include nut charcoal and animal charcoal. They are prepared in the same manner, except that the raw materials are nuts and bones, respectively.

Charcoal once occupied the position in the manufacture of iron that is now held by coke. In fact, coke was developed as a reducing agent to replace charcoal only when, shortly before 1600, it became evident in England that the use of the latter was making serious inroads on the forests of the country. Even today in the United States a little over 50 per cent of all the charcoal made is used by the metal industries. Household heating and cooking accounts for 35 per cent. The balance is used up in such fields as the manufacture of black powder, the making of various carbon products, and the preparation of stock and poultry feed.

Important by-products of charcoal manufacture include wood alcohol, acetic acid, combustible gases and acetone.

Liquid Fuels

The solid fuels just discussed contain carbon as the principal combustible substance. In some cases the carbon is free, while in others it is combined with hydrogen and oxygen. Most liquid and gaseous fuels, on the other hand, belong to a group of substances called *hydrocarbons*. These constitute the simplest class of organic com-

pounds and, as the name suggests, are compounds of carbon and hydrogen only. Several hundred hydrocarbons are known covering a wide range of molecule size. As an illustration of this variation in numbers of carbon and hydrogen atoms we will examine briefly the so-called *paraffin series* of hydrocarbons. This happens to be the series to which most of the hydrocarbons in liquid and gaseous fuels belong. The lower members are gases at ordinary temperatures, the simplest one being methane and having the formula CH_4. Methane has the lowest boiling point of the series, having to be cooled to —263° F. before it changes into a liquid. Progressing through the series to more complex molecules, one finds that the boiling points rise. Along in the middle of the group there are some which are liquids at normal temperatures. One of these, pentadecane ($C_{15}H_{32}$) boils at 520° F. and solidifies at 50° F., which indicates that it is a liquid at room temperature. At the upper end of the series are a few hydrocarbons which are ordinarily in the solid state, such as pentatriacontane. This compound with the awe-inspiring name has the formula $C_{35}H_{72}$ and melts at 167° F. Thus, in the same series, we have at one end molecules containing only 5 atoms and at the other, some with over 100 atoms in them. In this hydrocarbon series as well as in all others, as the number of atoms per molecule increases, the temperatures at which changes of state occur get higher. Although we are primarily interested in fuels in this chapter, it should be pointed out that not all hydrocarbons are fuels. Rubber, for one, is a hydrocarbon but quite obviously would not be a satisfactory substitute for coal for use in the furnace.

Petroleum

In its natural state petroleum is a liquid the color of which may be anything between light brown and black. It is a mixture rather than a compound. Mixed together in it are a large number of hydrocarbons together with some compounds of nitrogen, oxygen, and sulfur. Although we know it as a liquid, it contains hydrocarbons varying all the way from simple gases like methane up to solids containing 30 or more atoms per molecule.

In many respects the history of petroleum is like that of coal. The chief difference is that while coal's ancestry is traced back to land

vegetation, the organisms which form petroleum are believed to have been marine plants and animals.

The first oil deposits to be exploited in this country were those in Pennsylvania. At the present time these eastern fields furnish but a small fraction of the total crude oil produced in the United States. Approximately half of the world's supply now comes from California, Oklahoma, and Texas, with the United States as a whole producing some 63 per cent of the total.

Because of the great range of properties possessed by the various hydrocarbons which are mixed together in petroleum, no great use is made of it in its natural state. For this reason, practically all the petroleum produced is refined. The process of refining is primarily one of dividing up this conglomeration of many hydrocarbons in such a manner as to get together in small groups those which have about the same boiling points.

In other words, the basic refining process is fractional distillation, a procedure that was explained and illustrated in connection with the discussion of liquid air in Chapter 15. In what is known as " Batch Distillation " the petroleum is simply heated in large steel cylinders, the constituents vaporizing in order of increasing boiling points. The vapors are collected and condensed to give the separated products. In " Continuous Distillation " the process might more accurately be called one of fractional condensation. The crude petroleum is first conducted to an evaporator in which all constituents with boiling points lower than 750° F. are vaporized. This mixture of vapors is then slowly cooled, the constituents condensing out in order of decreasing boiling points. It should be noted that both of these are physical processes. No changes occur in the structure of the molecules themselves as they are sorted out according to the temperatures at which they vaporize.

Gasoline

Of the various fractions into which petroleum is divided, the group of hydrocarbons that we call gasoline is by far the most important to us today. Since our use of gasoline is almost entirely as a fuel for internal combustion engines, like those in automobiles and airplanes, it is the degree of volatility of the hydrocarbon which deter-

mines whether or not it belongs to this favored group. It so happens that the hydrocarbon molecules having from 5 to 10 carbon atoms fall within the proper range. These distill off after the volatile gases like methane which possess from 1 to 4 carbon atoms per molecule. It will be obvious from this that the volatility of any gasoline could be raised by including in it more of the smaller molecules, or lowered by increasing the number of the larger ones present. Gasoline companies make adjustments of this sort in the mixture for different seasons of the year.

Fractional distillation alone would be entirely inadequate as a means of providing all of the gasoline used today. Another method designed to supplement the distilling process and increase the amount of gasoline that could be got from a barrel was introduced as early as 1913. This process is called *cracking*. As the name suggests it is essentially a method by which hydrocarbon molecules containing too many carbon atoms to be within the gasoline range are cracked or broken up into smaller ones. This, of course, is a chemical process. The decomposition is made to occur by subjecting the heavier molecules to temperatures of from 750° F. to 900° F. Since at ordinary atmospheric pressure such high temperatures would simply cause the hydrocarbons to boil away, the process is carried out under pressures which sometimes are as high as 800 or 900 pounds per square inch.

In England and Germany, where large supplies of natural petroleum are not available, the manufacture of synthetic gasoline from low grades of coal has received considerable attention. The process which is used is known as *hydrogenation*. It was developed by Friedrich Bergius who was awarded the Nobel prize in 1931 for his work in this field. First, the coal is powdered. Then, under considerable pressure, it is heated to a fairly high temperature in an atmosphere of hydrogen. In the reactions which occur hydrogen unites with the carbon in the coal forming a black tarry liquid which is a mixture of hydrocarbons. This oil is similar to natural petroleum in makeup and can be refined in much the same manner. Although the use of coal in this way has not been important in the United States, the success of the process suggested that it might be tried on certain natural crude oils of low quality. It has been used in this fashion with considerable success and large yields of gasoline and oil obtained. The process oc-

curs with an expansion of volume to the extent that the volume of the finished product may be 103 per cent to 105 per cent that of the raw materials.

Kerosene

In the process of fractional distillation of petroleum the constituent that boils off next after gasoline is the material that we know as kerosene or coal oil. It contains 10 to 15 carbon atoms per molecule and is, therefore, somewhat less volatile than gasoline. Although it is used to a small extent in certain types of internal combustion engines, it is no longer enormously important as a fuel. At one time kerosene was the main product of petroleum refining, and the gasoline was thrown away because there was no use for it. The advent of the automobile reversed that situation so that today the kerosene is one of the by-products.

Other Liquid Fuels

Although so far man's ingenuity in finding new ways to get greater yields of desirable products from natural petroleum has about kept pace with his increasing use of this valuable liquid, it is conceivable that at some future time the supply might become seriously depleted. If such a situation should arise, there are other possibilities in the way of liquid fuels. Among these are the alcohols. Grain alcohol is made by the fermentation of sugar and so the sugar plantation offers one possible source of new fuel supplies. Wood alcohol, as has already been mentioned, is a by-product of the manufacture of charcoal.

Gaseous Fuels

The fact that most gaseous fuels, like those belonging to the liquid group, are hydrocarbons has already been mentioned. It is also true, as one would expect from the fact that boiling points rise with increased number of atoms per molecule, that the gaseous hydrocarbon molecules are simpler than those of the liquids.

Natural Gas

The mixture which we call natural gas consists almost entirely of hydrocarbons, with from 80 per cent to 90 per cent being methane

(CH_4), the gas already mentioned as the simplest one of the paraffin series. Methane is also known under two other names, *marsh gas* and *fire damp*. The first of these is applied to it when it is the product of decay of vegetation in swampy regions, and the second when the gas occurs in coal mines. In the latter case its presence is extremely dangerous since it is highly explosive when mixed with air.

Natural gas is usually found in porous sandstone or limestone which is covered with a dome-shaped layer of gas-tight rock. Usually associated with it and below it is petroleum or salt water or both. Natural gas has been found at depths varying from a few hundred feet to eight or nine thousand, and under pressures running all the way from a few pounds per square inch to as much as a ton. In order of their yield, the states which produce most of our natural gas are West Virginia, Oklahoma, Pennsylvania, Ohio, Louisiana, and California.

A somewhat recent development in connection with natural gas has been the storage of it in liquid form by cities for use in periods when demand exceeds supply. Such an emergency is particularly likely to occur in very cold weather. As a liquid at a temperature of —250° F., it can be stored in about one six hundredth of the space it occupies normally as a gas. This means that 7 or 8 quarts would supply the average home's needs, exclusive of heating, for one day. Stored in liquid form under pressure it can simply be allowed to expand into the gas mains and vaporize at times when the normal supply is low.

Coal Gas

The most common of the manufactured vapor fuels is coal gas. It has already been mentioned as a by-product of the manufacture of coke. Its composition varies somewhat with the coal used but in general combustible materials constituté about 97 per cent of its total mass. Of this a little less than half is methane and other hydrocarbons, with the remainder being mostly hydrogen. Some carbon monoxide is usually present also.

Water Gas

This fuel cannot be classed as a hydrocarbon since it consists of a mixture of carbon monoxide (CO) and hydrogen (H_2). Water gas

is made by passing steam through a bed of burning coke or hard coal. The reaction, which occurs at about 1,800° F., is described as follows:

$$\text{Carbon} + \text{Water} \rightarrow \text{Carbon monoxide} + \text{Hydrogen}$$
$$C + H_2O \rightarrow CO + H_2$$

The process, which also is one way of manufacturing hydrogen, is not continuous because of the fact that heat is absorbed during the reaction instead of being driven off. Consequently the temperature soon drops below the point necessary for the reaction to occur, at which time the steam is turned off and the fuel heated again by means of a forced draft.

Hydrogen and carbon monoxide are present in water gas in approximately equal parts and together make up about 93 per cent of the total. The 7 per cent of incombustibles consist largely of carbon dioxide and nitrogen.

Producer Gas

If, in the above process, the steam is replaced by air containing only a little steam, the product is called producer gas. Since air is about 78 per cent nitrogen, and nitrogen is incombustible, producer gas is only from 40 per cent to 45 per cent composed of substances that will burn. The principal one of these is carbon monoxide. This makes it a much poorer fuel than either water or coal gas in terms of heat units per cubic foot. Its cost of manufacture, however, is so low that it compares very favorably with the others on a basis of heat units per dollar.

Acetylene Gas

When lime and coke are heated together, one of the products is a material called calcium carbide (CaC_2), the reaction being as follows:

$$\text{Lime} + \text{Carbon} \rightarrow \text{Calcium carbide} + \text{Carbon monoxide}$$
$$CaO + 3C \rightarrow CaC_2 + CO$$

Then when water comes into contact with the calcium carbide the following reaction occurs:

$$\text{Calcium carbide} + \text{Water} \rightarrow \text{Calcium hydroxide} + \text{Acetylene}$$
$$CaC_2 + 2H_2O \rightarrow Ca(OH)_2 + C_2H_2$$

As will be evident from its formula, C_2H_2, acetylene is a hydrocarbon. It burns with an exceedingly bright flame and produces intense heat. It forms an explosive mixture with air over a wide range of mixture ratios and may even be exploded by shock. Today its principal use is in the oxyacetylene blowtorch with which temperatures as high as 6,000° F. are obtained. Before the use of electricity became so general, acetylene was employed in bicycle and automobile lamps and even in home lighting. In the latter case the fuel was stored under pressure in a large tank, usually buried underground for safety, and the gas conducted to the lighting fixtures through a system of hollow wires.

Questions

Discussion

1. State the relationship between the terms "combustion" and "oxidation."
2. Distinguish between endothermic and exothermic reactions.
3. Why is spontaneous combustion less likely to occur when oil-soaked rags are left out in the open than when they are crumpled up in the corner of an unventilated closet?
4. Look up some additional information about dust explosions.
5. Why should not a 100 per cent concentration of fuel gas burn better than a 15 per cent concentration?
6. Describe the process of coal formation mentioning the outstanding properties of each of the various kinds.
7. List at least four important by-products of coke manufacture.
8. Why may the common practice of burning soft coal for heating purposes be called a wasteful process?
9. Describe the process called "destructive distillation" and mention two products in whose manufacture it is involved.
10. What is the general relationship between the number of atoms in a hydrocarbon molecule and the state in which it exists (solid, liquid, or gas)?
11. Contrast the ancestry of a gallon of gasoline with that of a ton of coal.
12. Distinguish between "batch distillation" and "continuous distillation."
13. Distinguish between "cracking" and "hydrogenation" as applied in the gasoline industry.
14. Name three artificial fuel gases and describe their manufacture.

Multiple Choice

1. Coke is made by (a) reducing iron ore, (b) the oxidation of graphite, (c) the fractional distillation of charcoal, (d) passing steam over hot carbon, (e) heating soft coal in the absence of air.

2. List all of the following that are products of the process of heating wood out of contact with air. (a) Wood alcohol. (b) Coal tar. (c) Ammonium sulfate. (d) Acetic acid. (e) Charcoal.

3. Which one of the following fuels is neither a hydrocarbon nor a mixture of hydrocarbons? (a) Natural gas. (b) Acetylene. (c) Coal gas. (d) Water gas. (e) Gasoline.

4. Arrange the following fuels in order of increasing percentage of carbon content. (a) Bituminous coal. (b) Lignite. (c) Coke. (d) Peat. (e) Anthracite coal.

5. Which one of the following statements is true with respect to petroleum? (a) It is a by-product of coke manufacture. (b) It is refined by fractional distillation. (c) It represents the third step in the formation of coal. (d) It is a chemical compound. (e) It originated from land vegetation.

True–False

1. All four of the types of coal are found in the United States.
2. Lignite has a lower heat value per pound than does peat.
3. Rusting is an example of an endothermic reaction.
4. The average molecule of the hydrocarbons found in kerosene contains more carbon atoms than does the average hydrocarbon molecule in gasoline.
5. A liquid fuel that was sprayed into a chamber where the temperature was higher than the ignition point of the fuel might be expected to burn more rapidly than if it were poured in.
6. Peat can only begin to form in regions where the climate is dry enough to prevent decay.
7. In the manufacture of water gas, reduction occurs.
8. All oxidation reactions are exothermic.
9. From every ton of soft coal put in a coke oven about three-quarters of a ton of coke is obtained.
10. In the manufacture of gasoline by cracking, the hydrocarbons used as raw materials contain an average of fewer than five carbon atoms per molecule.

Chapter 17

METALS AND ALLOYS

ONE basis of classification of the elements, groups them into metals and non-metals. It is with the metals that we shall concern ourselves in this chapter, considering first the group as a whole.

GENERAL DISCUSSION OF METALS

Historical Background

As one might expect, the first metals to be used by primitive man were those that are found free in nature to the greatest extent. These are gold, silver, and copper. They seem to have been employed by the Egyptians in the manufacture of jewelry and dishes as early as 5000 B.C. Coins made from various combinations of these metals were also in use at an early time. The Old Testament makes several references to such coins being used by the Hebrews. Tin entered the metal picture when someone discovered, probably accidentally, that if it were mixed with copper the resulting substance was harder, and would take and hold an edge better than either constituent. So there came into being the alloy that we call bronze, a material which was so important in the ancient world that its name is given to one of the cultural stages in human development. The Bronze Age began in Egypt around 3000 B.C. and in Europe some 500 or 1,000 years later.

Since almost no iron exists free in nature, it undoubtedly came into general use somewhat later than those just mentioned. However, the Egyptians and Assyrians seem to have made some use of iron a number of centuries before the birth of Christ. Homer describes weight-throwing contests in Greece, in which a piece of iron served both as the object to be thrown and as the prize to be awarded the winner. Also various references are made to iron by Biblical writers. As soon as methods were developed for separating iron from its ores

in reasonably large quantities and at fairly low cost, it ceased to be classed as a precious material and began its career as the world's most valuable metal from the standpoint of actual use.

Lead, which is another metal mentioned in the Bible, was used in Rome in making water pipes. Greek writers refer to mercury and seem to have some knowledge of its power as a solvent for other metals.

By the time medieval alchemy reached its peak, gold, silver, copper, tin, iron, lead, and mercury were known as "The Seven Metals," and were linked in the speculative writings of the alchemists with the various other "sevens" which were so important to them, including the seven known heavenly bodies, the seven openings in the head, and the seven days of the week.

One other metal, zinc, seems to have been known long before the days of the alchemists in its role as one of the constituents of the alloy brass. Aristotle mentions brass in his writings, and reference is made to it in the Old Testament, although there is some evidence to indicate that early Biblical writers used the word somewhat indiscriminately to mean copper or any of its alloys. Almost no metals other than these seven or eight were known until the eighteenth century and many that we use today, not until the nineteenth.

Properties as a Group

It will be recalled that the physical properties of a substance are those which do not involve changing it into another material. These include melting and boiling points, hardness, conductivity, strength, toughness, and density. Of those which we naturally associate with metals, the outstanding one is metallic luster. All metals possess this characteristic, although in the case of those like potassium and sodium, which oxidize very rapidly, it is only noticeable on a freshly cut surface. Also, metals are mostly solids at ordinary temperatures and have comparatively high melting points, the exception being mercury. They are for the most part good conductors of heat and electricity, with silver being the best in these respects. They exhibit to a comparatively high degree the properties of being able to be drawn into fine wires and hammered into thin sheets, characteristics that are called ductility and malleability respectively. An ounce of gold can

be drawn into a wire almost 50 miles long or hammered into a sheet that has an area of between 175 and 200 square feet.

In considering the chemical properties of metals, the first point to be noted is that they vary widely in degree of chemical activity, some being enormously active and others very inert. Therefore the metals are listed in order of decreasing chemical activity in what is called the "Chemical Activity Series." The more common metals are given in Table 15 in the order in which they appear in this series. Hydrogen may seem to be a little out of place in the group since we do not ordinarily think of it as a metal. Chemically, however, it behaves in many respects like a metal and so is usually included in the activity series.

TABLE 15. THE CHEMICAL ACTIVITY SERIES

1. Potassium	(K)	9. Tin	(Sn)
2. Sodium	(Na)	10. Lead	(Pb)
3. Calcium	(Ca)	11. Hydrogen	(H)
4. Magnesium	(Mg)	12. Copper	(Cu)
5. Aluminum	(Al)	13. Mercury	(Hg)
6. Zinc	(Zn)	14. Silver	(Ag)
7. Iron	(Fe)	15. Platinum	(Pt)
8. Nickel	(Ni)	16. Gold	(Au)

Now, keeping this in mind, let us examine two or three chemical reactions which involve metals. The one most familiar to us is undoubtedly rusting or corrosion, which, we learned in Chapter 15, is simply a special case of oxidation, differing from combustion only in the speed with which the reaction occurs. We think of rusting primarily as it applies to iron, because we see more evidence of it with that metal than with any other. The fact that iron, the most used of all the metals, rusts at a comparatively rapid rate causes man to spend many thousands of dollars every year painting iron structures in order to keep the oxygen in the air from coming into contact with the metal. A glance at the above table shows that there are six metals above iron, that is, six that are more active and so should oxidize more rapidly. Of these, the two whose presence in this location is probably most surprising are aluminum and zinc, since we do not commonly think of these metals as corroding very rapidly. Actually, they do rust more readily than iron. But it happens that after the outer layer of these

metals oxidizes, it clings to the metal underneath and so acts as a protective covering that prevents oxygen from reaching the inner portion of the metal. Tin and lead behave in a similar fashion. With iron the rust flakes off about as rapidly as it forms and so continually exposes a fresh surface to the atmospheric oxygen. The metals toward the bottom of the series oxidize very slowly, and in the case of silver, platinum, and gold, almost not at all.

Potassium and sodium, the first two metals in the series, react with water very violently. In the case of the latter when the metal goes into solution and displaces hydrogen, the reaction occurs so vigorously that enough heat is generated to melt the sodium not yet dissolved. The fourth metal, magnesium, is scarcely acted upon at all by cold water although it does go into solution in boiling water. From aluminum, on through the rest of the series, there is little chemical reaction with water.

We will approach a third kind of reaction exhibited by metals, by asking how one might go about determining experimentally the order of their activity. A little thought will make it clear that neither of the two reactions just discussed would be likely to prove satisfactory. In the case of oxidation the fact that the rust or oxide clings to certain ones of the group would make it difficult to compare those particular metals with each other. Then it would be almost impossible to determine the sequence of the metals toward the bottom of the series, because they all exhibit such slight activity with both water and oxygen. Experiment has shown that one of the best ways of comparing chemical activities is by means of what is called a "displacement reaction." To illustrate the procedure, let us assume that iron and copper are to be compared. We may begin by immersing a piece of either metal in a solution of the other one. If we happen to choose first to dip a piece of copper in a solution such as iron sulfate ($FeSO_4$), we find that no reaction whatever occurs. But if we place an iron nail in copper sulfate ($CuSO_4$) a coating of copper immediately begins to form on the nail. Analysis shows that the following reaction is occurring:

$$\begin{array}{ccccccc} \text{Iron} & + & \text{Copper sulfate} & \rightarrow & \text{Copper} & + & \text{Iron sulfate} \\ \text{Fe} & + & \text{CuSO}_4 & \rightarrow & \text{Cu} & + & \text{FeSO}_4 \end{array}$$

The layman might be tempted to describe this phenomenon as a kind of chemical alienation of affections, in which the iron replaces the copper in the copper sulfate family. The chemist describes it by saying that iron displaces copper from solution. Since we found in the first part of the experiment that copper would not displace iron, we may say that iron is the more active of the two metals and so would be placed above copper in the chemical activity series. In this manner the comparative activities of all the metals can be determined.

The fact that the list of metals which we have called the " Chemical Activity Series " is also known as the " Electromotive Series " suggests one other point that may be mentioned here. If any two of these metals are used as the electrodes of an electric battery, the one nearer the top of the series will be the negative pole. The reasons for this will be considered in a later chapter.

Metallurgical Processes

The fact that very few metals occur naturally in the free state in any abundance has already been mentioned. In general, metals are found to be chemically combined with one or more other elements, the compounds being known as ores. Metallurgy may be defined as the science of extracting the metals from their ores. This entire separation process can be broken down into four general operations.

Concentration. — The first of the four is called concentration. It is true in the mining of almost all metals that only a small fraction of the material taken from the mine actually consists of the desired ore. Therefore, the first thing to do is to separate, as completely as possible, the ore from the rocks, soil, or other earthy material with which it is mixed. It is this separation that constitutes concentration. For the most part concentration processes simply accomplish more efficiently the result that would be achieved by having a large number of men go over, by hand, all the material removed from the mine and pick out the pieces of ore. In one of the oldest concentration procedures, the entire mixture is carried along by running water. Separation is accomplished by gravity, the heavy ore sinking to the bottom. This method is exemplified on a small scale in " panning " for gold.

One of the most important modern concentration processes is

called *oil froth flotation*, and is, in a sense, the reverse of the above. In this case the material that comes from the mine is ground up and thoroughly mixed with water and oil. The oil must be one which will wet or adhere to the ore but not to the other substances, in somewhat the same sense that water will wet glass but not paraffin. The oil drops then cling to the ore particles and carry them along when they rise to the surface of the water, making it possible for the ore to be removed while the impurities collect below. Magnetic ores are sometimes concentrated by making use of the fact that they can be attracted by magnets.

It will be noted that the processes mentioned above are all purely physical ones. There are cases, however, in which all the material that comes from the mine is treated with a liquid which will dissolve the metal but not the impurities. Thus by chemical means the metal is concentrated directly in a solution from which it can be recovered in any one of several ways.

Roasting. — This second metallurgical process is not necessary when the ore is a compound of the metal and oxygen, that is, an oxide. If, however, the ore is a sulfide, or a carbonate, or a compound with arsenic, it is usually changed over into an oxide by roasting. As the name implies, the process is one in which the ore is heated to a high temperature in an excess of air. Under these conditions, ores which are above copper in the activity series change into oxides.

Reduction. — We now have the ore in the form of a compound of the metal and oxygen, either because we found it that way in nature, or because of the roasting process. The next step is to free the metal from the oxygen, a procedure to which we have already given the name reduction. The reducing agent used and the details of the method employed depend on the metal. The way in which the process is carried out in the case of iron will be discussed at some length a little farther on in this chapter.

Refining Processes. — Finally, there remain certain refining processes before the metallurgy is completed. These include both the removal of impurities that remain in the metal after reduction of the ore, and the addition of various substances which will give the finished product the particular properties that are desired.

We have gone from the mine to the refined metal in three or four

types of operations. It should be remembered that these are quite general and that variations may exist in any given case. Now we will investigate some of the metals individually, discussing iron in some detail, and several others much more briefly.

Iron and Steel

Occurrence and Properties

If it is true that the core of the earth is largely composed of iron, then iron is by far the most abundant element in our planet. However, as was mentioned in Chapter 7, it ranks fourth in abundance in the only part of the earth to which man has access, that is, the lithosphere. The three elements present in greater percentages are oxygen, silicon, and aluminum.

Pure iron is almost never found in natural deposits. The only exceptions are certain meteorites and a few deposits in Greenland. There is some evidence to indicate that the latter are also of meteoric origin. As might be expected from the rapidity with which iron rusts, the world's most abundant iron ores are compounds of the metal and oxygen. The principal ore mined in the United States is hematite or ferric oxide (Fe_2O_3). Other iron ores which are of some importance, although not used to any great extent in this country, include magnetite (Fe_3O_4), bog iron ore ($Fe_2O_3(H_2O)_3$), and siderite ($FeCO_3$).

The United States is the most favored of all nations in the extent and quality of its iron ore deposits now being mined. Brazil is believed to have as rich or richer ones which are too inaccessible at the present time to be worked profitably. France, as of 1939, ranks second in the magnitude of its iron production. Other valuable deposits occur in Germany, Sweden, Russia, China, and Morocco. More than 85 per cent of the iron used in this country comes from the Lake Superior region.

In its pure state, iron is silvery white, has a density between seven and eight times that of water, and is rather highly ductile and malleable. It cannot be tempered and is too soft to be of much practical value. Rather peculiarly, iron rusts much less rapidly when it is pure, than when it contains the usual impurities, although the impurities by themselves are less active chemically than is iron.

Reduction of Iron Ore

Concentration of iron ore is accomplished by one or more of the methods previously mentioned. Since the ore is already an oxide, roasting is unnecessary. Therefore we will proceed immediately to the consideration of the reduction or smelting procedure. In the case of

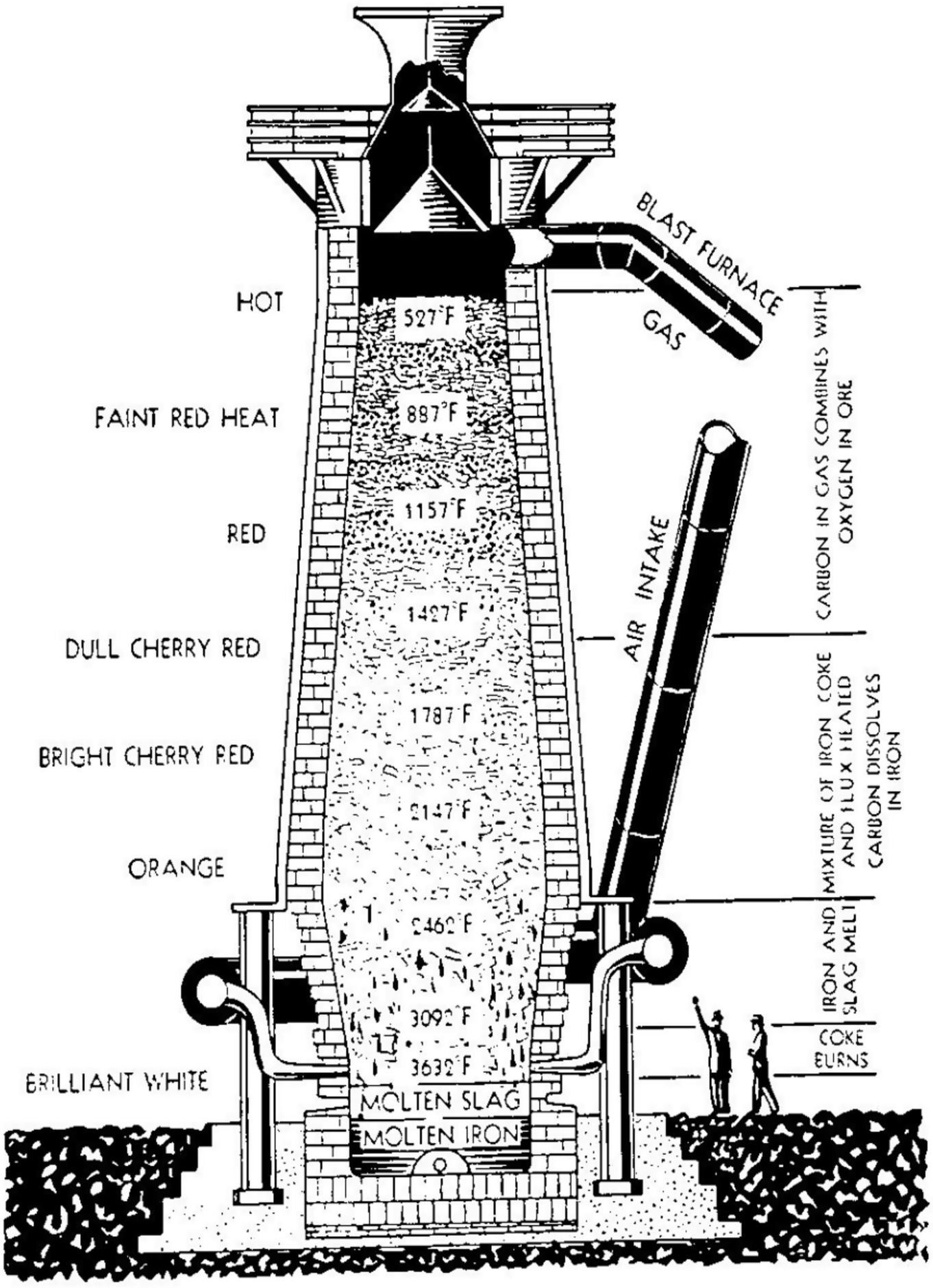


FIG. 89. Cross-sectional view of a blast furnace. (General Motors Corporation.)

iron ore this is done in the blast furnace, a drawing of which is shown in Fig. 89. The furnace itself may be as much as 90 feet in height and have a maximum diameter of 25 feet. The tapering part of the furnace above the widest point is called the "bosh" and the lower part the "crucible." The construction is of steel plates lined with firebrick. Once started the furnace is kept in operation continuously until forced to shut down for repairs, which may be a matter of months or even several years. There are four substances which may

be called the raw materials that go into the blast furnace. Three of these are solids, namely, iron ore, coke, and limestone. They are put into the furnace at the top through a kind of double trap door arrangement, in the proportion of 8 tons of ore to 4 tons of coke to 1 ton of limestone. The double doors prevent the escape of gaseous prod-

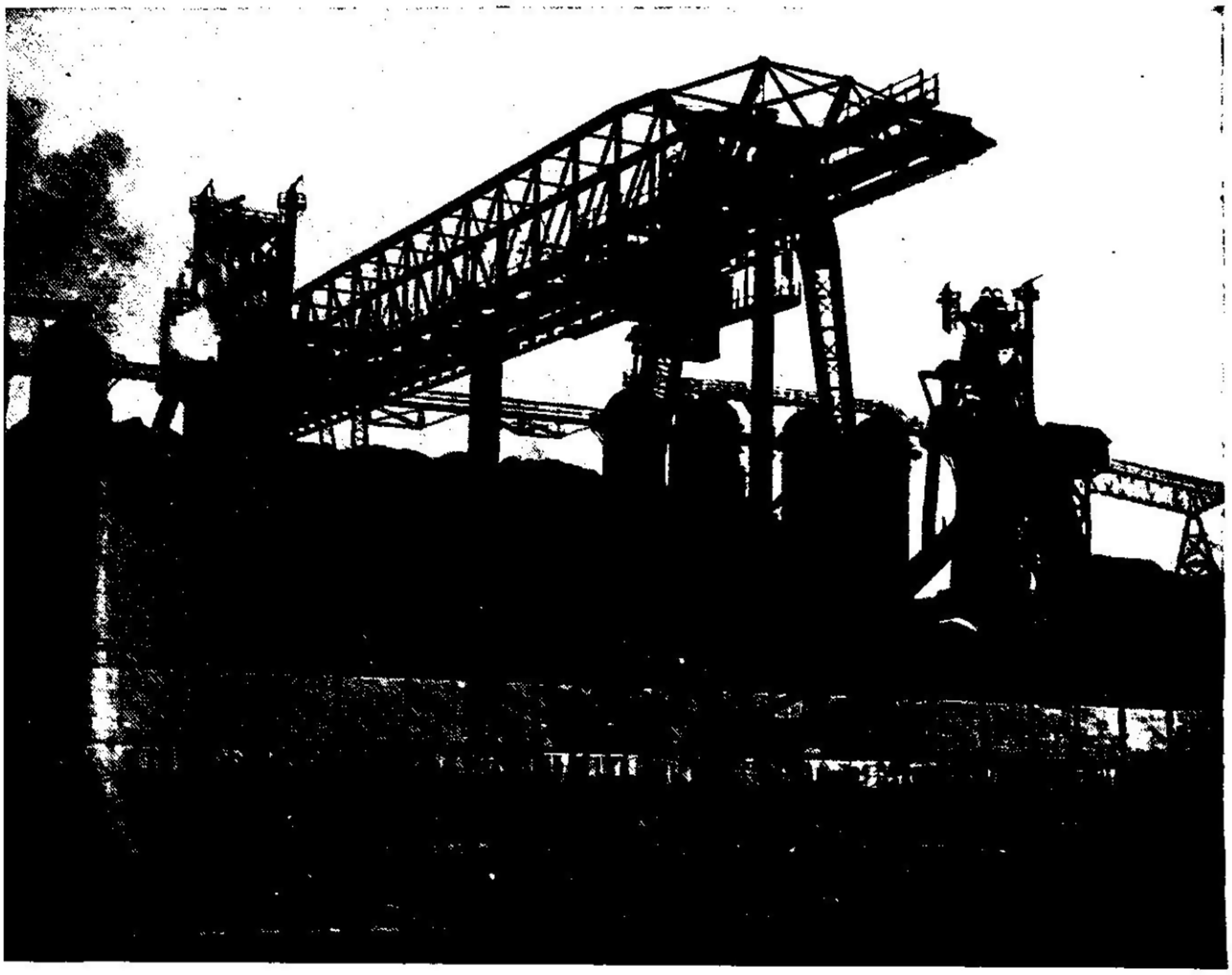


FIG. 90. Here iron ore becomes pig iron. The blast furnaces proper are the tallest of the structures in the background. The shorter cylinders are stoves for heating the air used in the blast. A pile of ore and the ore bridge for carrying the ore to the blast furnace are shown in the foreground. (American Iron and Steel Institute.)

ucts while the furnace is being charged. The fourth raw material is air that has been heated to a high temperature in stoves which adjoin the blast furnace. This air is conveyed to the furnace by the pipe marked " Air Intake " in the picture, and introduced below the solid charge. A complete unit consisting of a blast furnace and four stoves is shown in Fig. 90.

The reactions that occur in the blast furnace are principally of two kinds. First, we will describe the ones directly connected with the reduction of the iron ore. The coke at the bottom burns vigor-

ously in the abundance of hot air, generating temperatures as high as 3,600° F. This results in the formation of carbon dioxide,

$$\text{Carbon} + \text{Oxygen} \rightarrow \text{Carbon dioxide}$$
$$C + O_2 \rightarrow CO_2$$

As the carbon dioxide filters up through the coke above it, it unites with more carbon, forming carbon monoxide

$$\text{Carbon} + \text{Carbon dioxide} \rightarrow \text{Carbon monoxide}$$
$$C + CO_2 \rightarrow 2CO$$

The carbon monoxide then acts as the agent which reduces the iron ore, the reaction being,

$$\text{Iron Ore} + \text{Carbon monoxide} \rightarrow \text{Iron} + \text{Carbon dioxide}$$

For hematite: $Fe_2O_3 + 3CO \rightarrow 2Fe + 3CO_2$ or

For magnetite: $Fe_3O_4 + 4CO \rightarrow 3Fe + 4CO_2$

The molten iron drips down and collects at the bottom of the crucible. The hole called the " iron notch " is opened every few hours and the iron allowed to flow out into forms in which it hardens, becoming the material that we call *pig iron*.

Iron ore always contains a number of substances in addition to iron and oxygen. Among these are some silicon compounds which, for the most part, melt readily at blast furnace temperatures and so flow to the bottom of the furnace. One of the most common ones, however, is silica (SiO_2) which does not melt easily and so must be converted into another form. The limestone which was mentioned as one of the raw materials causes this change to occur. First the limestone decomposes into lime and carbon dioxide, as follows,

$$\text{Limestone} \rightarrow \text{Lime} + \text{Carbon dioxide}$$
$$CaCO_3 \rightarrow CaO + CO_2$$

The lime then reacts in the following manner with the silica to form calcium silicate, a compound that melts quite readily.

$$\text{Lime} + \text{Silica} \rightarrow \text{Calcium silicate}$$
$$CaO + SiO_2 \rightarrow CaSiO_3$$

All the molten siliceous compounds collect at the bottom of the furnace and float on top of the liquid iron. They are drained off through

an opening called the " cinder notch," and when solidified compose the material which we call slag, a substance that has commercial uses in road building.

A third product of the blast furnace is the flue gas which escapes at the top through a pipe called a " Downcomer." It is about one-fourth carbon monoxide and so has considerable heating value. Before being used, it is cleaned in a " Dust Catcher." Then about one-third of it goes to the stoves where it is burned to pre-heat the air entering the blast furnace. The stoves are large steel towers filled with a checker work of firebrick. The cycle of operation in any one stove is the following. Flue gas is burned until the brickwork is heated to a high temperature. Then the gas is turned off and air blown through the stove to be heated on its way to the blast furnace. When the stove cools down, the air is turned off and the stove heated again with flue gas. In order to permit continuous operation each furnace is accompanied by four stoves. At any given time, two of them are in the process of being heated while the other two are furnishing air to the furnace. A large blast furnace is capable of producing as much as 1,000 tons of pig iron per day.

Pig iron contains from 2 per cent to 5 per cent of carbon, several per cent of silicon and smaller amounts of sulfur, phosphorus, and manganese. Because of its carbon content it is brittle and of low tensile strength, although very hard. Its hardness, combined with the fact that it expands slightly when it solidifies, gives it value as a material from which to cast certain objects. Cast iron is used in stoves, radiators, wheels, and pipes. For applications requiring toughness, ductility, and greater tensile strength, the pig iron must be refined.

Refining Processes

Manufacture of Wrought Iron. — Before the days of low cost steel, wrought iron was used almost exclusively where toughness and great tensile strength were required. The device in which wrought iron is made is called a " puddling furnace." It consists of a sort of hearth covered with a low curved roof, which serves to reflect heat from an outside fire down upon the charge within. The materials put into the furnace are pig iron and some iron oxide. The former melts

and the latter oxidizes many of the impurities in the iron. Among the oxidation products are gases which escape, and liquids which remain mixed with the iron as slag. It happens that as iron is purified its melting point rises and so as the treatment continues the molten material eventually becomes a pasty mass, which is then rolled into a huge ball and removed from the furnace. The slag is squeezed out and the material that remains is called wrought iron. Since it contains less than half a per cent of carbon, it has lost most of its brittleness. It is tough, has a high tensile strength, and can be readily forged and welded. In earlier times it found its principal uses in the manufacture of such articles as chains, anchors, bolts, nails, and wire. Today it has largely been replaced by soft steel.

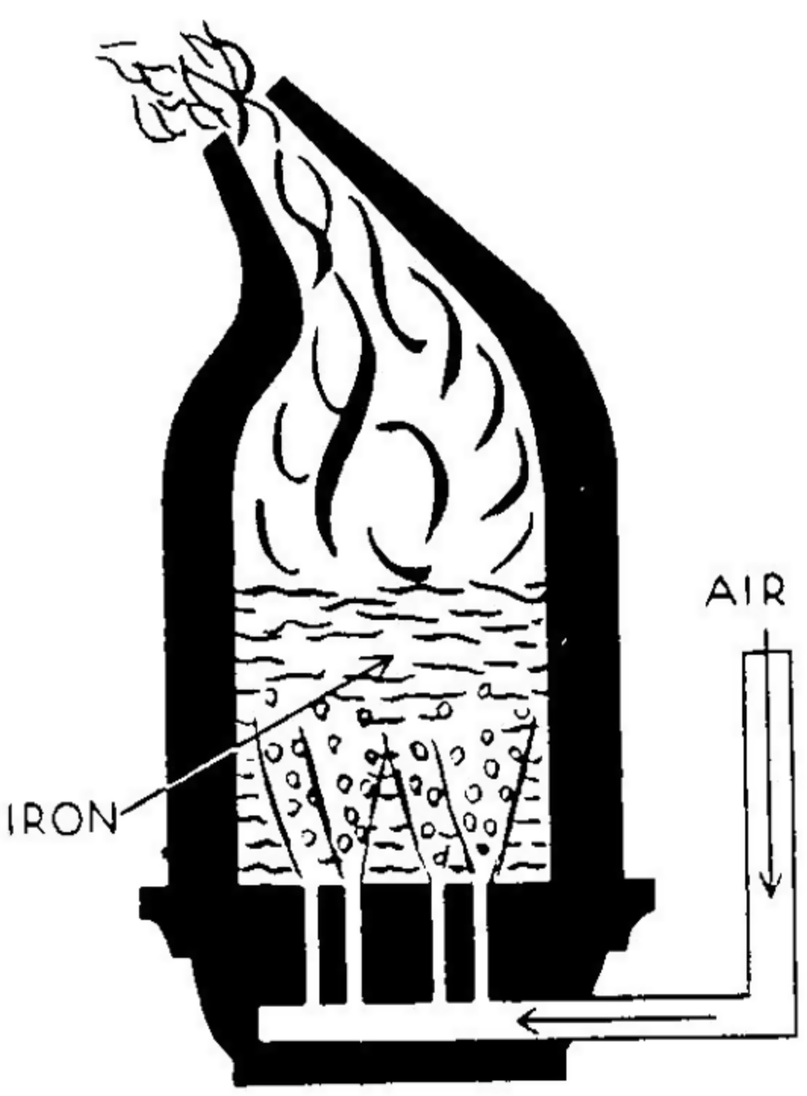


FIG. 91. Cross-sectional view of Bessemer converter. (After General Motors pamphlet, *Metallurgy and Wheels.*)

Manufacture of Steel. — The carbon content of steel varies from .1 per cent to 1 per cent. For this reason it avoids both the extreme softness of pure iron and the great hardness and brittleness of pig iron. The several manufacturing processes by which steel is made have in common the following general procedure: the first step is to remove, as far as possible, all impurities; then the next step is to put in the materials necessary to give the finished product the desired properties.

The manufacturing process which first made cheap steel possible, was developed by the Englishman, Sir Henry Bessemer, and introduced in 1855. The device in which Bessemer steel is made is called a Bessemer converter. As Figs. 91 and 92 indicate, the converter is a pear-shaped furnace mounted so that it can be rotated about a horizontal axis. The furnace itself is some 18 feet high and 12 feet in diameter. It is loaded with from 15 to 20 tons of molten pig iron. The so-called " blow " consists in forcing a stream of air up through the molten metal. The air oxidizes the various impurities in the iron, and in so doing, generates large amounts of heat. The flames die

down after something like 15 minutes and the " blow " is over. Before removing the purified iron, exact amounts of carbon, manganese and other elements are added to give the finished steel the desired group of characteristics. The steel is then poured into a huge ladle in which it is carried to molds. Bessemer steel is a comparatively low cost product since one converter can produce 15 to 20 tons every 10 or 15 minutes. However, this rapidity of production has the disadvantage of making accurate control of conditions difficult. Because of this the product is not of the highest quality or uniformity. Bessemer steel finds its principal applications in railroad rails, machinery castings, and construction work.

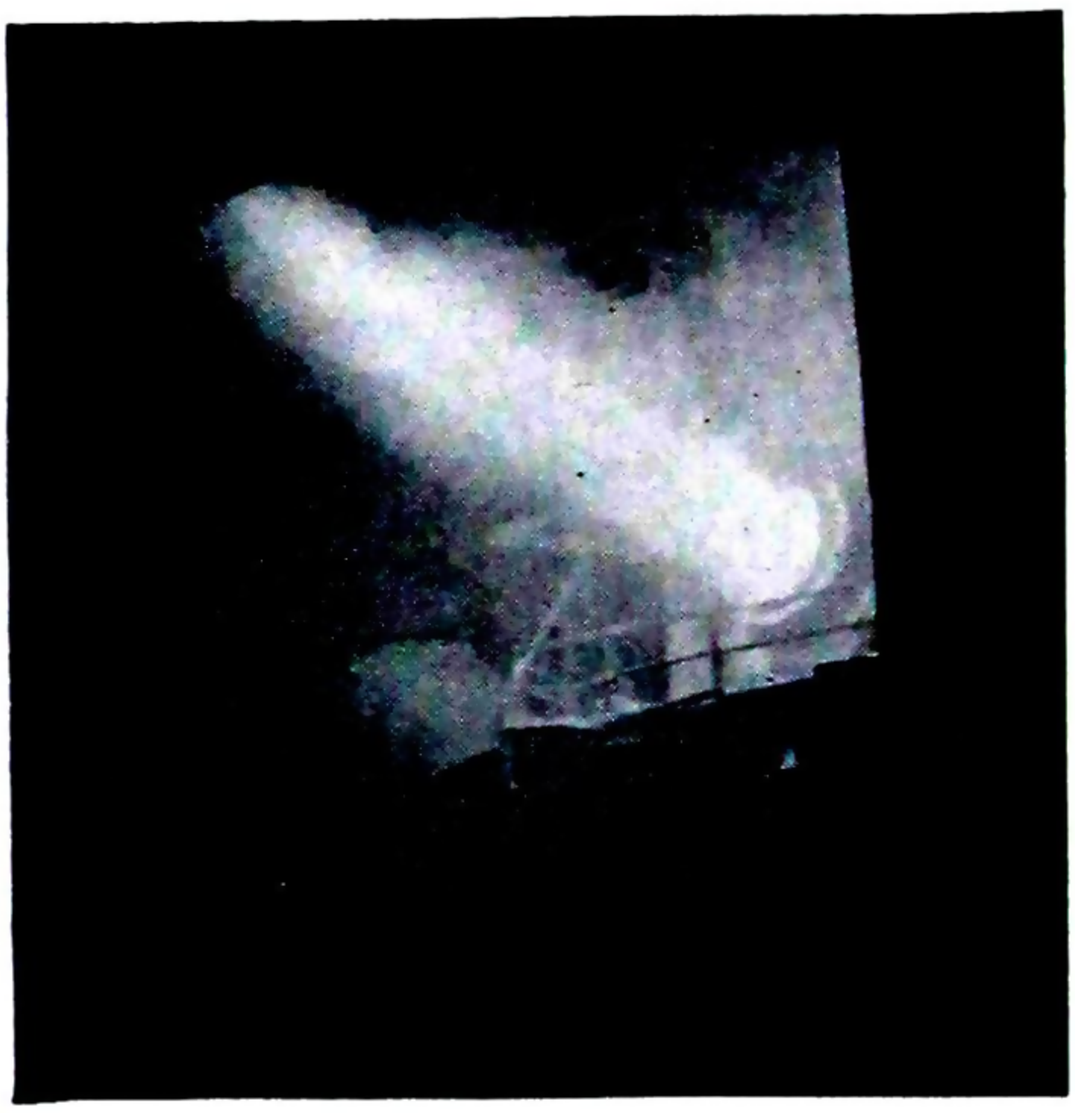


FIG. 92. Air bubbling through molten pig iron in the Bessemer converter refines the pig iron into steel. The long tongue of flame emerging from the mouth of the converter is caused by the burning out of the carbon in the pig iron. (American Iron and Steel Institute.)

A better grade of steel is made by the "open hearth process." An open hearth furnace is shown in Fig. 93. The hearth area is rather large but the furnace is not very high. It takes a charge of from 50 to 100 tons of molten pig iron. Air that has been heated in the brick checker-work along the sides flows over the iron and is reflected down upon it, burning out the impurities. Better control of the procedure is possible than in the previous case because of the fact that something like 8 hours are required for each charge of pig iron. After the impurities have been removed and while the iron is still in the furnace, the desired amounts of carbon and other substances are added. Over three-fourths of the steel made today is made in the open hearth furnace. One of its chief advantages is the fact that fairly high percentages of scrap iron and scrap steel can be included with the pig iron that goes into it. This is not true of the Bessemer converter.

Very high-grade steel, such as is required for tools, automobile and airplane parts, and in similar applications, is made in an electric furnace. Here very high temperatures can be attained as well as a maximum of control over temperature during the steel-making process. The electric furnace is usually employed to further refine Bessemer or open hearth steel rather than to make steel directly from pig iron.

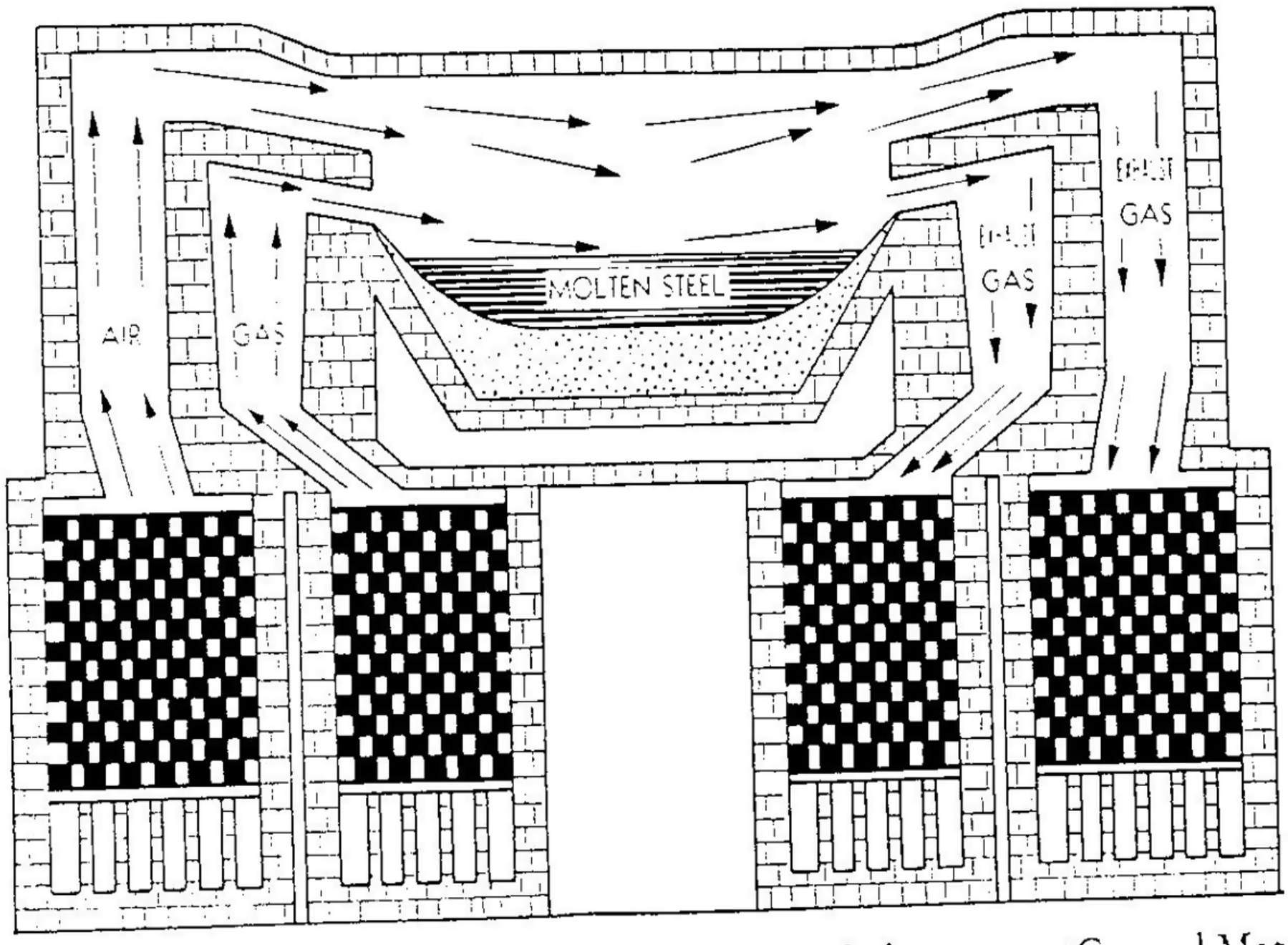


FIG. 93. Cross-sectional diagram of an open hearth furnace. (General Motors Corporation.)

The development of these processes has caused the annual production of steel in the United States to increase from less than a million tons to around 50 million tons during the past 60 years. During the same period the price of a ton of steel has decreased from about $160 to approximately $60.

SOME OTHER METALS

Copper

We have already mentioned the fact that copper was one of the first metals to be used by man, the reason being that it is found free in nature to a comparatively large extent. In the light of our discussion of chemical activity the reader may infer from this that copper is a relatively inactive metal. Most of the copper used today comes from compound ores, although an appreciable amount is still taken

from deposits of the native metal. The most valuable copper deposits in the United States are in northern Michigan. Most of the compound ores contain sulfur or oxygen; for example, chalcocite (Cu_2S), chalcopyrite ($CuFeS_2$), malachite ($Cu_2(OH)_2CO_3$), and cuprite (Cu_2O). These differ so widely in composition it is not surprising that a number of different metallurgical processes are employed. After concentration the oxide ores are, in general, reduced with coke and a flux much as is iron. The sulfur ores, which account for about three-fourths of this country's copper output, must be roasted before they can be reduced. Very pure copper, such as is required for electrical conductors, is made by treating the products of the above operations electrolytically. This process is described in some detail in Chapter 25 in connection with copper and silver plating. It will be sufficient at this point to say that it is similar to electrolysis of water in which water is decomposed into hydrogen and oxygen by passing an electric current through it. In the case of copper, the impure metal is used as one of the electrodes. The electric current then causes this metal to go into solution. It travels through the liquid and comes out of solution at the other electrode, being plated on it in the form of the pure metal.

Most people are quite familiar with many of the uses to which copper is put. Its value lies in its properties of high ductility and malleability, high conductivity of both heat and electricity, and low rate of corrosion or oxidation. Most of the copper produced today goes into alloys, of which the following are some typical examples: gold jewelry, of gold and copper; silver jewelry, of silver and copper; silver coins, of silver and copper; five-cent pieces, of nickel and copper; bronze, of copper and tin; and brass, of copper and zinc.

Aluminum

Aluminum is the most abundant metal, and the third most abundant element in the earth's crust, being present to an average extent of about 8 per cent. It achieved no commercial importance before 1886 because it is not found free in nature and up to that time no cheap method of separating it from its ore was known. It was first isolated in 1825 and within a few years of that date small amounts of it were produced at a cost of approximately $160 per pound. About

1854, Napoleon III became interested in the metal as a substance from which to make armor. He authorized a government grant of funds to be used in searching for better and cheaper methods of freeing aluminum from its natural compounds. This work, directed by the French chemist Deville, was carried on for over 30 years and by 1886 had reduced the cost of manufacture to $12 per pound. Following the discovery in 1886 of a cheap method of extracting aluminum the price dropped rapidly and by 1888 reached $2 per pound. The present peace-time cost of a pound of aluminum is in the neighborhood of 20 to 25 cents.

The discovery of the method now in use for getting aluminum in a pure form constitutes, in many respects, one of the most interesting and astonishing stories in all science. Perhaps the most startling feature of all in connection with it is the fact that the discovery was made quite independently at almost exactly the same time, by Charles Hall, a 22-year-old American and by Paul Herolt a 22-year-old Frenchman. The two men received their patents the same year, each in his own country. Neither had any connection with, or knowledge of, the other up to the time of their remarkable discovery and so deserve equal credit for it. Hall first became interested in preparing pure aluminum while he was a student in Oberlin College. It occurred to him to try to apply electrolytic processes to the problem. Since such methods require that the substance be in liquid form, it was necessary to find something in which bauxite (Al_2O_3), the aluminum oxide ore, would dissolve. The answer to this problem proved to be a mineral called cryolite, the composition of which is indicated by its formula, Na_3AlF_6. Hall found that the cryolite could be melted easily and then the aluminum oxide ore simply dissolved in it. Then by passing an electric current through the solution, he discovered that the ore would separate into its constituents. Thus, less than a year after he graduated from college, Charles Martin Hall solved one of the most baffling chemical problems of his day.

The enormous importance of aluminum in our modern world is due to the group of highly desirable properties which it possesses. The fact that, when properly treated, it has a high tensile strength although just slightly more than a third as heavy as steel, makes it a valuable construction material for use in modern streamline trains

and airplanes. As an insulating material, aluminum foil is as efficient as cork and weighs less than one-fiftieth as much. The use of aluminum pigment instead of lead in the paint put on a battle cruiser saves some 100,000 pounds in the boat's weight. The fact that both lightness and good heat conducting properties are combined in aluminum makes it valuable for use in automobile and airplane engine parts such as pistons, and cylinder heads. These same two properties, plus its pleasing appearance, give it an important place in the field of cooking utensils. Because it is a comparatively efficient conductor of electricity, steel reinforced aluminum electrical transmission lines are in use in the United States today to an extent of between 400,000 and 500,000 miles. Compared with copper in this respect, aluminum is a poorer conductor for the same diameter and length wire but a better one for the same weight of transmission line. It is constantly finding new uses in the architectural field. The first important application in this field occurred in 1884 when an aluminum cap was put on the Washington monument. Its architectural importance today may be gauged by the fact that some 3,000,000 pounds of it went into Rockefeller Center.

Zinc

Since the most important zinc ore in the United States is a sulfide (ZnS), roasting enters into its metallurgy. When the zinc sulfide is roasted at low temperatures, it changes into the sulfate,

$$\text{Zinc sulfide} + \text{Oxygen} \rightarrow \text{Zinc sulfate}$$
$$\text{ZnS} + 2\text{O}_2 \rightarrow \text{ZnSO}_4$$

The zinc sulfate is then dissolved in sulfuric acid and the free metal recovered from this solution by electrolysis. Zinc is a bluish-white metal a little lighter than iron. It is hard and brittle at ordinary temperatures but becomes quite malleable and ductile if heated a little higher than the temperature of boiling water. Its principal use is in coating iron to prevent rusting of the latter. Such zinc-coated iron is the material that we call " galvanized iron." It should be recalled that although zinc appears above iron in the activity series, its corrosion is negligible because of the surface layer of oxide that forms first and protects the rest of the metal. Its next most important use is as the

material which is alloyed with copper to make brass. It also finds application in such fields as the manufacture of dry batteries, paint pigments, and roofing. An important use of the compound zinc oxide, which will be mentioned again in the next chapter, is as a compounding ingredient in the manufacture of rubber products.

Tin

The principal tin ore is the oxide SnO_2, known as cassiterite. The ore is reduced to the metal by the same process used with iron, that is, by smelting with carbon. Tin is a white metal, not quite so heavy as iron. At the temperature of boiling water it is highly malleable but becomes very brittle if heated to 400° F. At low temperatures it tends to crumble into dust, a phenomenon which is known as " tin disease " and which occurs most rapidly around —50° F.

Much of the tin used in this country goes into making our annual supply of some 12 million " tin cans," which actually are iron covered with a very thin coating of tin. It also forms a number of valuable alloys including bronze, which has already been mentioned, and solder. The annual production of tin is in the neighborhood of 130,000 tons. Very little is produced in the United States although we use about two-thirds of the world's total production.

Lead

The lead ore of greatest importance is the sulfide PbS, called galena. In extracting the lead, part of the lead sulfide is first roasted to produce an oxide. This oxide is then mixed in a blast furnace with the sulfide ore, limestone and coke. In the smelting process the lead is set free from both compounds.

Lead is about half again as heavy as iron and is very soft. Its tensile strength is low. It has a bright luster when freshly cut but quickly tarnishes, forming a coating which protects the rest of the metal. It is rather highly resistant to a number of acids. Some of its applications are as a constituent of type metal, solder, and Babbitt metal; in storage batteries, electric fuses, collapsible tubes like tooth-paste containers, lead pipes, lead foil, and other products. Some of its compounds find important uses in the paint, glass, rubber, match, storage battery, linoleum, and textile industries.

The above group comprise the better known and most important metals. Analogous facts about others may be found in chemistry and metallurgical textbooks. Some of the less familiar ones which are rapidly increasing in importance because of their alloys with each other and with the more common ones are nickel, tungsten, chromium, vanadium, molybdenum, and cobalt.

Alloys

Alloys are usually prepared, simply by melting two or more metals together and then allowing this liquid mixture to cool and solidify. If no chemical reaction occurs between or among the constituents, the resulting alloy is a simple mixture. In some cases there is a definite reaction and the resulting alloy is a chemical compound. In a third type the degree of combination of the metals is such that the product can be described best by calling it a solid solution.

The enormous importance of alloys lies in the fact that by combining metals in this fashion almost any desired set of properties can be attained. In any specific application, one decides first what group of characteristics would be most desirable and then, in most cases, finds it possible to alloy several metals together and produce just the required material. The field of magnetic alloys offers striking examples of this fact. In certain situations it is desirable to have a material which can be very highly magnetized and which will retain this magnetism for long periods of time. For such applications an alloy of aluminum, nickel, cobalt, and iron has been developed. A magnet made of this substance, called Alnico, will support several hundred times its own weight. A magnet with exactly the opposite set of properties would be one which could be very easily magnetized and which would lose its magnetism just as readily. An alloy of nickel and iron, known as Permalloy, is outstanding in these respects. It acquires magnetism so easily that a bar of it held north and south, that is, parallel to the earth's magnetic field, will pick up small bits of permalloy. Also it loses this magnetism so readily that if the bar is turned east and west, the pieces of permalloy fall off, being no longer attracted. Both of these alloys are mentioned somewhat further in Chapter 23.

Other examples of designing an alloy to fit the use for which it

is intended are found in the field of low-melting-point materials. Wood's metal, an alloy of bismuth, lead, tin, and cadmium, melts at 160° F., a temperature considerably lower than the boiling point of water, and lower than the melting point of any of its constituents. One important application of this material is in automatic sprinklers, such as are used in buildings for fire protection. When a fire starts the heat melts out a plug of Wood's metal rather quickly, and releases the water. In 1936, Professor Sidney French of Colgate University discovered that by adding the somewhat rare metal indium to the four in Wood's metal, an alloy was obtained that has a melting point of 116° F., which is less than 20 degrees above body temperature. Various interesting uses for this material have been suggested. Used as casts for broken bones it would be hard at room temperature and yet could be melted off when the fracture was healed simply by immersing the arm or leg in fairly warm water. In casting art objects, a model made of this material could be plated with a hard metal, and then the alloy melted and poured out, leaving an accurate and durable mold. Seamless hollow vessels could be made in the same manner.

Type metal, composed of antimony and tin, is another example of an alloy which has a melting point lower than that of either of its constituents. Although the melting points of antimony and tin respectively are about 630° C. and 327° C., a mixture of 10 per cent of the former and 90 per cent of the latter melts at approximately 230° C.

Among the steel alloys important advances have been made along several lines. The addition of manganese or nickel improves the strength of steel. Molybdenum steels are harder and more heat resistant than the ordinary product. The addition of chromium and a small amount of nickel produces stainless steel, which is highly resistant to corrosion as well as being as strong and tough as the older forms. Alloying vanadium with steel makes it less subject to fatigue under repeated shock.

Space does not permit the discussion of any more specific alloys The following table lists a number of the more important ones together with their percentage composition and applications.

TABLE 16. A PARTIAL LIST OF IMPORTANT ALLOYS

Name	*Per Cent Composition*		*Uses*
Aluminum brass	Copper	68–70	Propeller blades, rudder frames, sea valves
	Zinc	27–31	
	Aluminum	1–3	
Brass	Copper	73–66	Sheets, tubes, cartridges, etc.
	Zinc	27–34	
Bronze	Copper	88	Strong valves and fittings
	Tin	8	
	Zinc	2	
	Lead	2	
Aluminum bronze	Aluminum	10	Hard, non-corrodible, ornamental
	Copper	90	
Babbit metal	Tin	70–90	Bearings
	Lead	7–24	
	Copper	2–3	
Bell metal	Copper	80–75	Bells, gongs
	Tin	20–25	
Chromel	Nickel	60	Resistance wire for electric heating (Patented)
	Chromium	40	
Duralumin	Aluminum	95.5	Strongest and best aluminum alloy. Airplane and automobile parts
	Copper	3	
	Manganese	1	
	Magnesium	.5	
Wood's metal	Bismuth	38	Automatic sprinklers, fuse plugs
	Lead	31	
	Tin	15	
	Cadmium	16	
German silver	Zinc	21–30	Table ware, cheap jewelry
	Nickel	18–25	
	Copper	(remainder)	
Gold (white)	Gold	82	Jewelry
	Nickel	18	
Invar	Iron	64	Precision instruments, clocks, etc. Has low coefficient of expansion
	Nickel	36	

Name	*Per Cent Composition*		*Uses*
Magnalium	Aluminum	90–94	Scientific instruments, balance beams
	Magnesium	10–6	
Manganese bronze	Copper	56	Propeller blades; has great resistance to wear
	Zinc	41	
	Tin	.5	
	Iron 1, Manganese .5, Aluminum 1		
Monel metal	Nickel	60	Propeller blades, wire, sheets, etc. Is almost non-corrodible
	Copper	33	
	Iron		
	Manganese	(?)	
Pewter	Tin	85–90	Platters, bowls, vases, etc.
	Lead	15–10	
Solder	Lead	67	Used by plumbers
	Tin	33	
Steel — low carbon	Carbon	.05–.25	Boiler plate, rivets, bridges, shafting, structural work
	Iron	(remainder)	
Steel — medium carbon	Carbon	.4–.8	Cutlery, rails, drills, castings, saws, files, razors, etc.
	Iron	(remainder)	
Steel — high carbon	Carbon	1.2–1.5	
	Iron	(remainder)	
Type metal	Lead	60–85	Printer's type
	Antimony	8–20	
	Tin	5–35	
Tin foil	Tin	88	Wrapping for foodstuffs, tobacco, chewing gum, etc.
	Lead	8	
	Copper	4	
Tinsel	Tin	60	Decorative purposes
	Lead	40	
Sterling silver	Silver	80	Silverware, etc.
	Zinc	18	
	Copper	2	
Silver coins	Silver	90	
	Copper	10	

Name	*Per Cent Composition*		*Uses*
Shot metal	Lead	99	Bullets and small shot
	Arsenic	1	
Stainless steel	Chronium	12–15	Uses requiring material that will not corrode
	Carbon	.3–.4	
	Manganese	.2–.5	
	Silicon	.2–.3	
	Iron	(remainder)	
Permalloy	Nickel	78	Telephone cables
	Iron	21	
	Cobalt	.5	

QUESTIONS

Discussion

1. Give reasons for gold, silver, and copper having been the earliest metals used by man.

2. Name the members of the group known to the alchemists as "The Seven Metals."

3. Define ductility and malleability.

4. Explain the "displacement reaction" method of determining comparative chemical activities of different metals.

5. Why are not corrosion and reaction with water satisfactory bases for comparing the chemical activities of all the metals?

6. Find out at least three specific displacement reactions other than the one given in this chapter.

7. Name and define the four general metallurgical processes.

8. Which one of the four metallurgical procedures is not necessary in the case of ores which are oxides? Explain.

9. What compound of iron is the most common ore in the United States?

10. Using equations, describe the reactions occurring in a blast furnace that are directly connected with the reduction of the ore.

11. What is the function of the limestone which is put in the blast furnace along with the ore and coke?

12. Name three products of the blast furnace.

13. Describe two methods of making pig iron into steel.

14. In what ways is the metallurgy of the following metals different from that of iron and in what ways similar?

(a) Copper. (b) Aluminum. (c) Zinc. (d) Lead.

15. Name the several types of alloys and distinguish between them.

16. Mention several specific alloys and show what advantages they posses over pure metals.

Multiple Choice

1. The third, eighth, and thirteenth metals in the chemical activity series are calcium, tin, and silver respectively; from this, one may infer that (one or more) (a) calcium is the most active of all the metals, (b) silver will be displaced from solution by either tin or calcium, (c) silver is more active than tin, (d) tin is less active than calcium but more active than silver, (e) calcium will displace either tin or silver from solution.

2. List all of the following that occur in the blast furnace in the manufacture of iron. (a) Production of steel. (b) Formation of slag. (c) Reduction of iron ore. (d) Production of wrought iron. (e) Alloying of materials to make a product like stainless steel.

3. Which one of the following does not belong in the group? (a) Hydrogenation. (b) Roasting. (c) Concentration. (d) Reduction. (e) Alloying.

4. Gold may be drawn out into very fine wire because of its high (a) density, (b) malleability, (c) ductility, (d) electrical conductivity, (e) tensile strength.

5. An open hearth furnace uses which one of the following as its principal raw material and produces which one? (a) Wrought iron. (b) Slag. (c) Steel. (d) Combustible gas. (e) Oil froth. (f) Pig iron. (g) Bessemer steel.

True–False

1. If tin (Sn) appears in the activity series above copper (Cu), one would expect the following reaction to occur:

$$Cu + SnSO_4 \rightarrow CuSO_4 + Sn$$

2. In the metallurgy of aluminum, reduction is accomplished chiefly by heating the metal in the presence of hydrogen.

3. The metals appearing near the top of the activity series were those used first by man.

4. The principal zinc ore requires roasting.

5. The Bessemer converter finds its principal application in the conversion of sulfur compounds into oxygen compounds.

6. No alloy can have a melting point lower than that of any of its constituents.

7. Some alloys are simple mixtures.

8. The history of tin, more than that of any other metal, would involve the use of the words " bauxite," " electrolysis," and Charles Hall.

9. A piece of any given metal will entirely oxidize when exposed to air, less quickly than any one above it in the activity series.

10. Stainless steel contains chromium.

Chapter 18

COLLOIDS AND RUBBER

RUBBER was chosen as one of the specific materials to consider in the "Man and Materials" portion of this book for two reasons. First, in its natural state it is typical of a group of substances called *colloids*. These, as was suggested in a previous chapter, constitute a state of matter that is immensely important to man. Then, second, rubber is well worth considering purely for its own sake. In it are combined a peculiar and valuable group of properties which make its many applications in our modern world both important and interesting.

THE COLLOIDAL STATE OF MATTER

It will be recalled that in Chapter 13 solids, liquids, and gases were referred to as "states of matter." As will be evident from the following discussion, this designation is used in a slightly different sense when applied to colloids. The general nature of these materials can be described best by the use of several illustrations.

If the reader throws a handful of marbles into a bucket of water they sink to the bottom almost immediately. Breaking the marbles up into smaller pieces increases the length of time required for all the material to settle. If this process of decreasing the particle size is continued indefinitely, a point will finally be reached at which no settling whatever occurs, even over a very long period of time. In such a case, the material is said to be in the colloidal state or to be a colloid. Synonymous terms for this state of matter are colloidal suspension and dispersion.

Size of Colloidal Particles

Before discussing the size of the particles involved in colloidal suspensions we must define a new unit of length, since the dimensions

which we will be considering are extremely small. The unit referred to is the micron. The word indicates that this unit is equal to a millionth of some other one, the other one in this case being the meter. It may also be defined as a thousandth of a millimeter which is approximately the equivalent of 1/25,000 of an inch. The micron is usually designated by the Greek letter μ. So we may say,

$$1 \text{ micron} = 1\ \mu = .001 \text{ mm.} = \frac{1}{25{,}000} \text{ inch}$$

In Table 17 the approximate size limits of various combinations of particles in a liquid are given in microns, millimeters, and inches.

When particles are smaller than those listed within the colloidal range, they are in the realm of molecular and atomic dimensions.

It should be emphasized that the limits between the groups mentioned in the table are not definite. Originally colloids were so named because their apparent solutions behaved like water solutions of glue in that they would not diffuse through certain kinds of membrane, the word " colloid " coming from the Greek word for glue (Κολλα). On the basis of later experiments it was concluded the peculiar behaviour of these " solutions " was due to the size of the suspended particles, and the colloidal range was defined as having the limits specified in Table 17. Today, however, systems which contain suspended particles larger than .5 μ are called colloidal if they exhibit the properties generally associated with colloids. These properties are discussed in the next section of this chapter.

Types of Colloidal Suspensions

The example discussed above of ground-up marbles in water illustrates but one of eight different types of colloids. In addition to a solid suspended in a liquid, there are various other possible combinations of solids, liquids, and gases. In fact, with but one exception, each one of these three can be suspended or dispersed in a medium like itself, or like either of the others. The exception is that of a gas in another gas, a combination which is not called a colloid because both parts, or phases, are of molecular size. Table 18 lists the eight possibilities together with one or two examples of each.

TABLE 17. COMPARISON OF PARTICLE SIZE IN VARIOUS SUSPENSIONS OF A SOLID IN A LIQUID

	Massive Particles	*Coarse Suspension*	*Cloudy Suspension*	*Colloidal Range*
Greater than	.5 mm. *or* 500 μ *or* .1 in.	.5 mm. to .005 mm. *or* 500 μ to 5 μ *or* .1 in. to .000,2 in.	.005 mm. to .000,5 mm. *or* 5 μ to .5 μ *or* .000,2 in. to .000,02 in.	.000,5 mm. to .000,000,1 mm. *or* .5 μ to .001 μ *or* .000,02 in. to .000,000,04 in.

TABLE 18. TYPES OF COLLOIDS WITH EXAMPLES OF EACH

1. Solid in a solid — cobalt oxide in glass — blue glass
2. Solid in a liquid — india ink
3. Solid in a gas — carbon in air — smoke; volcanic dust
4. Liquid in a solid — water in butter fat
5. Liquid in a liquid — mayonnaise or any emulsion
6. Liquid in a gas — fog; mist
7. Gas in a solid — floating soap
8. Gas in a liquid — foam

Properties of Colloidal Suspensions

There is a rather definite set of properties that all colloids possess. The first of the group, one which has already been mentioned, may be expressed by the statement, "*Colloids are stable.*" In other words, the material in suspension does not settle. Although gold has a much higher density than water, colloidal suspensions of gold in water may be kept for years without any of the solid collecting in the bottom of the container. The same is true of iron and other metals. Reasons for this stability will be suggested in connection with certain of the other colloidal properties.

One of them is implied in the fact that "*Colloids exhibit Brownian movements.*" Observed through an ordinary microscope, colloidal suspensions appear as clear and homogeneous as true solutions. This is because most colloidal particles have diameters less than the .5 micron which marks the lower limit of microscopic visibility. However, a device called the ultramicroscope makes it possible to detect the presence and movements of particles as small as a millionth of a centimeter which is but .01 micron.

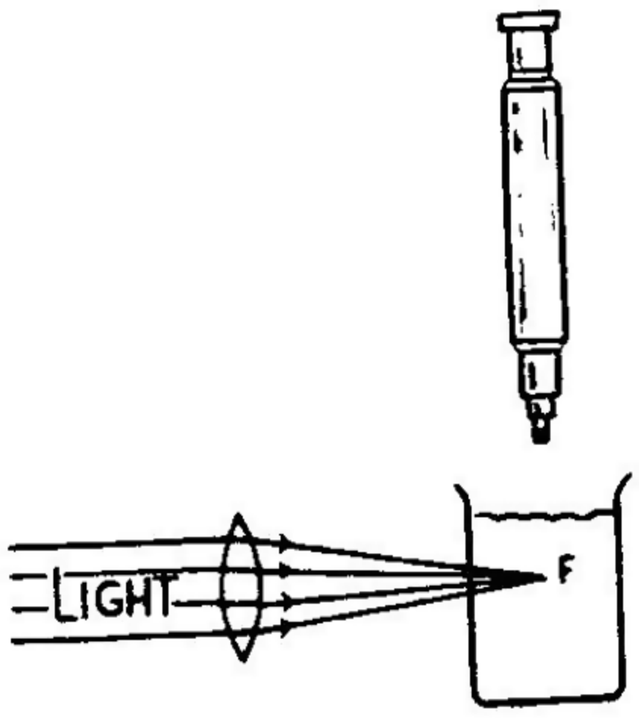


FIG. 94. Diagram illustrating arrangement of parts in the ultra-microscope.

Figure 94 illustrates the arrangement of parts in this instrument. A bright beam of light is brought to a focus at some point, *F*, within the colloid. If a high-power microscope is now directed at right angles to the direction of the beam of light, and focused on *F*, the colloidal particles will appear as tiny bright points of light on a dark background. When observed in this fashion, colloidal particles are seen to dart about

in exactly the same haphazard, zig-zag fashion as that described in our discussion of Brownian movements in Chapter 14. It will be recalled that this random movement was explained as being caused by the bombardment of the particles by molecules of the liquid. Such a general and continuous bombardment would tend to keep suspended particles from settling.

To distinguish colloids from true solutions, use is made of the fact that "*Colloids exhibit the Tyndall effect.*" When a beam of light is projected through a colloid its path may be clearly followed because of the reflection of light by the suspended particles. It is this phenomenon that is given the name "Tyndall Effect." The cone-shaped beam of light in the material is called a Tyndall cone. The effect is shown in the picture in Fig. 95. If the liquid used in that demonstration had been a true solution, the path of the beam of light would be quite invisible. A familiar example of the Tyndall effect is found in the appearance of a beam of sunlight coming through a hole in the window shade into a darkened room. Reflection of the light by dust particles in the air makes its path easy to follow. It is for this same reason that the beam of light from the projector to the screen in a motion picture theater is visible. Readers who were movie fans before the days of air-conditioned theaters will recall that the projector beam was much more conspicuous then than it is now.

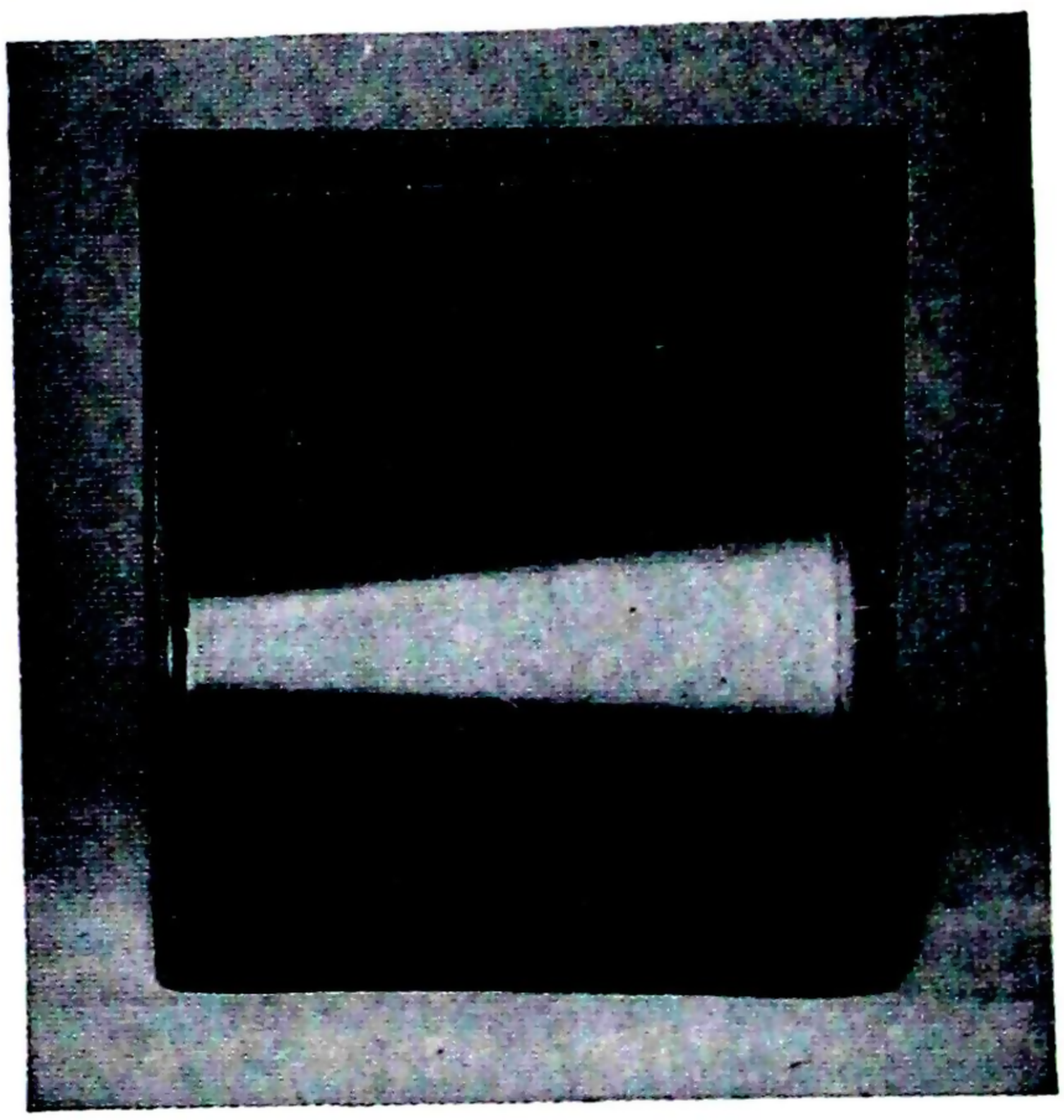


FIG. 95. Tyndall cone of light in a colloidal suspension. (Photo H. Walker.)

The last of the properties that we shall mention here is the fact that "*Dispersed colloidal particles carry electric charges.*" Anticipating the later discussion of electricity in this book, we may say here

that there are two kinds of electric charges, called positive and negative, respectively. Experiment shows that when two particles carry the same kind of charge they tend to repel each other. If they carry opposite charges, there is a force of attraction between them. Now in any given colloid all the particles have the same kind of charge, although different colloids may be charged differently. Thus there is always present a mutual repulsion within the dispersion. This

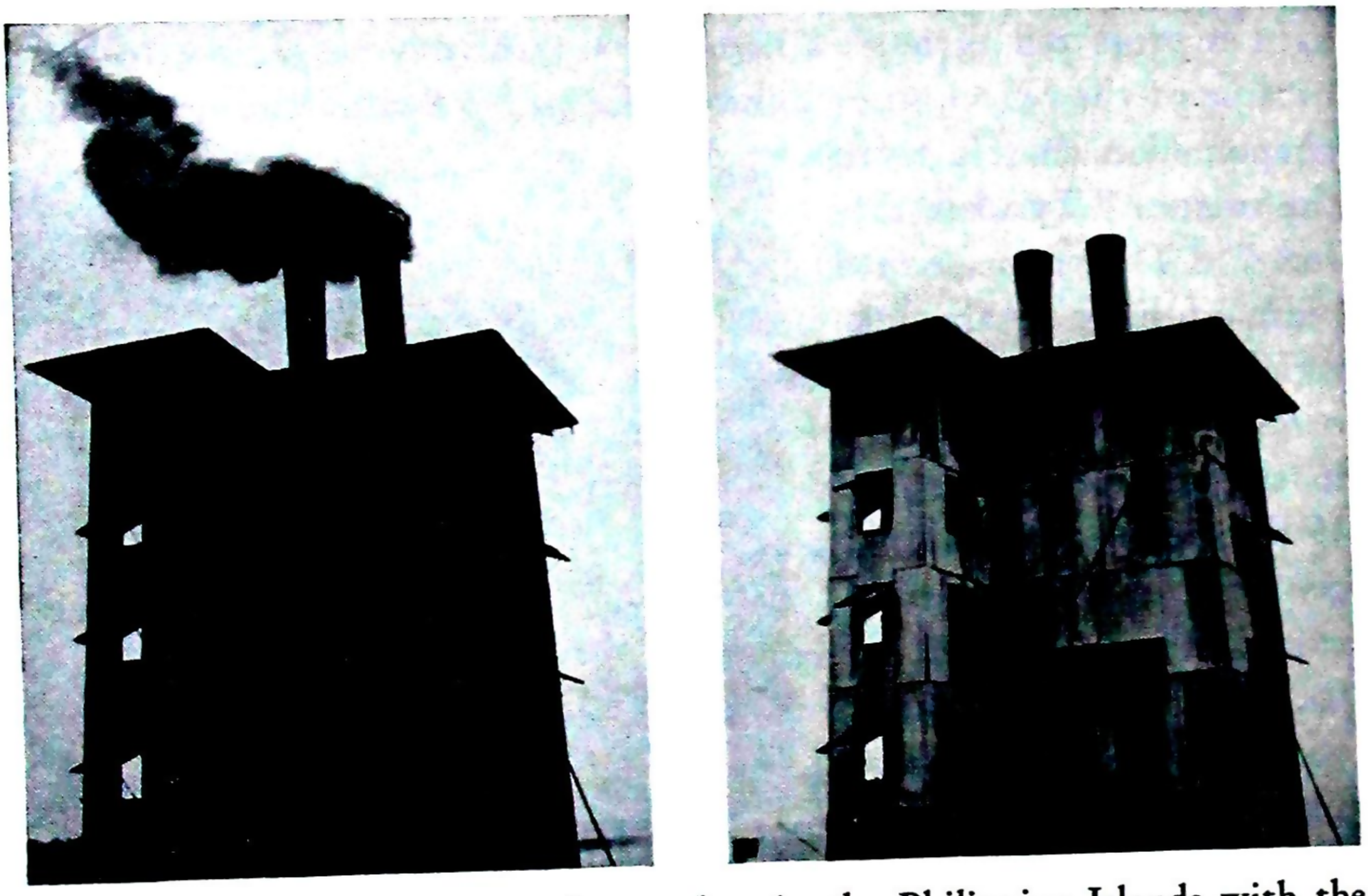


FIG. 96. Comparative views of a smelter in the Philippine Islands with the electric precipitator off and on. (Western Precipitation Corporation.)

tends to keep the small particles from either settling or coagulating into larger particles that might be heavy enough to sink to the bottom.

An industrial application of this last property is found in a method of eliminating factory smoke, known as "Cottrell Precipitation." Smoke is a colloidal suspension of carbon in air and the carbon particles are electrically charged. If the smoke passes over or between plates which carry the other kind of charge, the bits of carbon will be attracted. Once they reach the plate and are neutralized, they coalesce into larger particles which collect at the bottom of the chimney. Figure 96 shows comparative photographs taken of a factory with and without the Cottrell Precipitator in operation.

Having discussed colloids as a group, we next turn to a consideration of rubber which is colloidal in nature as it comes from the tree in the form of latex and remains so when it is prepared for the manufacturing processes and when it is made into articles of commerce.

RUBBER

Sources of Crude Rubber

Although there are more than 200 different trees, shrubs, and vines from which rubber can be obtained, not more than 10 or 12 are of any commercial importance. These are natives of different parts of the world and were all experimented with extensively during the early years of the rubber industry. The best source of rubber is the plant which possesses to the greatest degree such properties as high yield, high quality of product, early yield after planting, rapid healing of tapping wound, and ability to stand frequent tapping. The tree that has been found to excel in these attributes is Hevea brasiliensis which, as the name indicates, is a native of Brazil. The Hevea tree, as it is commonly known, like all rubber-producing plants requires a warm climate and an annual rainfall of 70 inches or more. It grows to heights of from 60 to 100 feet and averages between 3 and 4 feet in diameter. The latex is not, as is sometimes supposed, the sap of the tree. It flows up in canal cells which are located in the inner bark just outside the cambium layer. Since the fluid is under pressure in these cells, it will flow out when they are cut, and may be collected in a cup hung beneath the wound. The trees grow rapidly and reach bearing size in from 7 to 10 years after planting. At this time they are usually some 6 inches in diameter.

In the very early days of the rubber industry, all of the product was wild and practically all of it came from South America. With the growth of interest in Europe and North America, a period of experimentation began, during which trees and shrubs from essentially the entire world were tested for their rubber-bearing qualities. The net result of these investigations served to demonstrate the great superiority of the Hevea tree over all the others. Consequently, when extensive plantations were started, it was chosen as the tree to be planted, and today is responsible for some 97 per cent of the world's total yield of rubber.

In 1941 there were over 9,000,000 acres of plantations in production, averaging about 100 trees to the acre. Most of this acreage is in the Far East including Sumatra, Java, and Ceylon. The ancestry of every one of these hundreds of millions of Hevea trees on plantations can be traced back to some 70,000 seeds that were taken out of Brazil in 1876 by an Englishman named Henry A. Wickham. They were planted in the Kew Botanical Gardens and between 2,000 and 2,500 of them germinated. These were sent to Ceylon where cultivation was continued, with later distribution being made to other parts of the empire. Thus, there began the great eastern plantation system which increased its production from a total of four long tons of rubber in 1900 to almost a million and a half tons in 1940.

TABLE 19. WORLD RUBBER STATISTICS — ALL PRODUCTION IN LONG TONS (Part of table in *Rubber and Its Uses*, by H. L. Fisher, Chemical Publishing Company, 1941)

Year	*Plantation Acres*	*Eastern Plantation (Tons)*	*Amazon Valley (Tons)*	*Others (Tons)*	*Total (Tons)*	*U. S. Imports (Tons)*	*U. S. Imports % Total*	*Average N. Y. Price (per lb.)*
1826	.....	.....	.....	.....	.....	8	..	....
1830	.....	.....	.....	.....	.....	160	..	....
1840	.....	.....	.....	.....	.....	260	..	....
1860	.....	.....	.....	.....	.....	1,675	..	$0.62
1870	.....	.....	.....	.....	8,500	3,500	41	.87
1880	.....	.....	.....	.....	17,000	8,109	48	.85
1890	300	.....	.....	.....	30,750	15,336	50	.84
1900	7,000	4	26,323	17,804	44,131	20,308	46	.98
1910	1,125,000	7,269	37,938	48,806	94,103	42,210	45	2.07
1920	4,500,000	305,106	23,216	13,672	341,994	249,521	73	.36
1930	9,000,000	802,082	14,260	5,572	821,914	457,423	56	.10
1935	9,250,000	855,038	11,275	6,100	827,413	455,756	52	.12
1937	.....	1,108,717	15,576	11,105	1,135,398	592,528	52	.19
1939	9,000,000	970,029	13,892	18,847	1,002,768	486,372	49	.18
1940	.....	1,390,469	17,661	21,933	1,430,063	818,102	59	.20

Some world rubber statistics are shown in Table 19. The following points should be noted particularly. Plantation rubber rapidly surpassed the wild product in quantity and by 1941 made up over 97 per cent of the total production. It will be seen that ever since 1870 the United States has imported approximately half or more of the total and in 1940 took 59 per cent. The very considerable variation in price of rubber from year to year is worth noting. The high point occurred in 1910 when for a short time the price reached more than

three dollars per pound. In 1932, not shown in the table, rubber prices reached an all-time low of less than three cents per pound.

As soon as the War started in Europe, the consumption of rubber began to increase due to sales of war goods to England and France. In 1940, 648,000 tons were consumed and in 1941, 775,000 tons. At the end of 1940 the reserve stockpile of rubber in this country was only 125,800 tons, barely more than two months' supply. In the face of the threat of war in the Pacific which might cut off 90 per cent of our rubber supply, this country was compelled to seek the solutions to two problems. The first was the accumulation of as large a stockpile of rubber as possible, and the second, the development of a source of rubber independent of foreign supply.

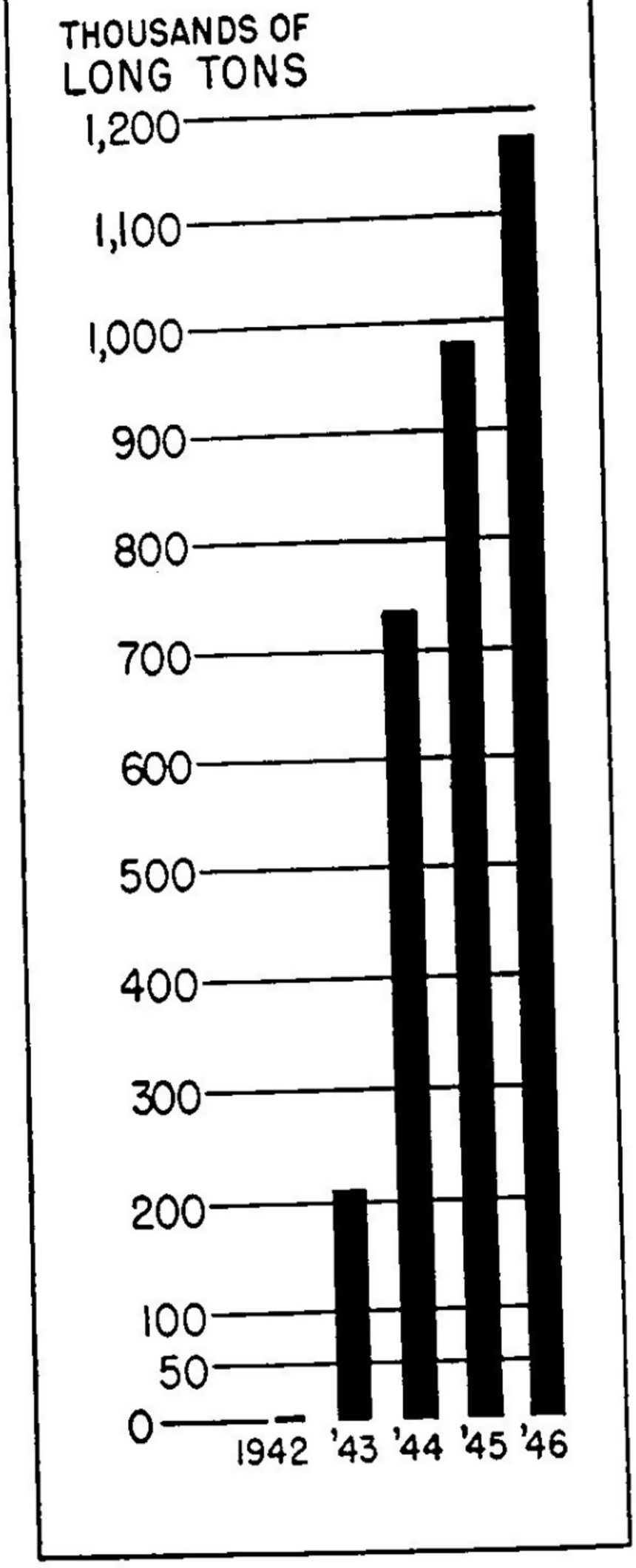


FIG. 97. Production of synthetic rubber in U. S. government-owned plants. (Rubber Age, July, 1945.)

The Rubber Reserve Corporation was organized to solve these two problems. As a result of the efforts of this organization, 1,835,000 long tons of natural rubber were imported during 1940 and 1941. By April 2, 1942, we had accumulated a reserve of 630,356 tons. If the reader will refer to a history of World War II, the significance of that date will be obvious. From this point on it was a race between the consumption of that stockpile and the building of factories to produce synthetic rubber to take its place. The charts in Fig. 97 show how this race progressed. The production shown for 1946 was that expected if the war had continued that long. At present (January, 1946) there is no way of knowing just how much natural rubber will be available this year. Fighting is still going on in Java, and in Malaya the process of replacing machinery and of

recruiting workers has just begun. By the time these pages are read, however, reconversion to natural rubber will probably be well along.

One of the most important developments of recent years in plantation rubber has been the return to the western hemisphere of the descendants of some of Wickham's 70,000 Hevea emigrants. The fact has already been mentioned that, although the United States uses over half the world's production of rubber, practically all the plantations of the world are in the Far East. This is true in spite of the fact that trees grown on these plantations are all natives of the western hemisphere. A number of projects are now under way looking toward the return of this particular native to its homeland.

A comparison of the yield of the trees being brought back to this hemisphere with that of the wild Hevea trees still growing in Brazil affords a striking example of the improvements that are possible through scientific selection and propagation. The wild jungle trees average around 3 pounds of rubber per tree per year while some fairly large cultivated acreages have attained an average annual yield of 10 pounds per tree. In a few cases trees have been developed to the point of producing as much as 24 pounds of rubber in one year.

Nature and Properties of Crude Rubber

As the latex comes from the wound in the Hevea tree, it is a colloidal suspension of rubber particles in a liquid which is almost entirely water. The rubber makes up something like 35 per cent by weight of the total. The suspended particles carry a negative electric charge, are roughly pear-shaped, and have an average diameter of approximately a quarter of a micron. It has been estimated that there are some 10^{14} of them per cubic inch of latex.

Chemically, crude rubber is classed as a hydrocarbon. Although the ratio of carbon atoms to hydrogen atoms in the rubber molecule is known, the exact number of these constituent atoms is not established. Therefore the chemical formula for rubber is written, $(C_5H_8)_x$ where x is believed to have a value of 1,000 or more. This formula simply means that the evidence indicates the rubber molecule to be a kind of chain composed of 1,000 or more links, each one of which is a C_5H_8 group. The molecular weight of rubber is very high, being

68,080 for $x = 1{,}000$, as compared with a little over 18 for water, 58.5 for table salt (NaCl), or 98.1 for sulfuric acid (H_2SO_4).

Two important chemical reactions involving rubber will be considered here. The first is the reaction between rubber and sulfur to which we give the name vulcanization. The discovery of vulcanization by Charles Goodyear in 1839 stands out as the most important single development in the history of this material. This is because of the fact, already referred to, that rubber by itself becomes soft and sticky when warm, and stiff and brittle when cold. When reacted with sulfur, that is, when vulcanized, it maintains its desirable properties over a considerable temperature range. Vulcanization techniques will be described in more detail in connection with the later discussion of steps in processing rubber.

The second reaction is oxidation. Crude rubber exhibits some tendency toward aging or oxidation, particularly when in sunlight or at fairly high temperatures. Unfortunately, after the rubber is vulcanized the rate at which this deteriorating reaction goes on increases. Consequently, it becomes necessary to add certain ingredients to the rubber to retard oxidation. Further mention of substances which serve this purpose will be made a little later in the chapter.

The best known physical property of rubber is undoubtedly its great extensibility. With the possible exception of certain muscles, rubber possesses this characteristic to a greater degree than does any other natural substance. In some forms it can be stretched as much as 1,000 per cent and still return to almost exactly its original size.

So-called true, or crystalline, solids have sharply defined melting points. For example, ice that is heated from 0° F. up to 32° F. remains a hard, solid material until the latter temperature is reached and at that temperature, melts. Other substances that we commonly call solids, such as glass and paraffin, simply grow progressively softer as they are heated and eventually flow. Technically, such materials are not solids but are super-cooled liquids. Rubber resembles these latter substances in that it does not have a definite melting point. It differs from them in that liquid rubber formed by heating is chemically different from the unheated rubber.

Two other important physical properties of rubber are its high resistance to abrasion or frictional wear and its ability to withstand

shock. Both of these are of great importance in certain applications that will be mentioned in the latter part of the chapter.

Among rubber's less desirable characteristics is its poor resistance to liquids like oil, gasoline, or benzene. Contact with these causes it to swell and to lose its strength. It is in applications involving such contact that some of the new synthetic elastics find important uses.

Finally, it may be mentioned that although rubber shows poor resistance to the liquids cited above, it is almost entirely inactive in the presence of many chemicals, which accounts for various important modern applications of it.

Steps in Rubber Processing

It is our purpose in this section to trace briefly the various stages in the treatment of rubber from the time it oozes from the Hevea tree as latex to the time it becomes a finished product.

Plantation Operations. — We will begin this life history of a rubber tire or hot-water bottle with the tapping of the tree. This process is pictured in Fig. 98. The slanting cut is made about a quarter of an inch deep and extends from a third to a half of the way around the tree. The process requires considerable skill because, to be completely satisfactory, the cut should extend across all the latex-carrying cells but should not be deep enough to injure the cambium layer. The exuding latex flows along the cut to a spout which directs it into a small cup. In successive tappings the bottom of the original cut is lowered. This also is evident in the picture. A healthy, mature tree can be tapped on an average of every other day for many years, yielding from 1 to 2 ounces of latex each time. In the morning the native tapper taps from 350 to 400 trees. By the time this is finished the flow from the first one is completed and he makes the rounds again collecting the latex, which is taken to a central station.

Here the rubber must be coagulated, that is, separated from the water in which it is suspended in the latex. In the early days of the rubber industry this was accomplished by evaporation. A native would permit the latex to drip slowly on a horizontal stick which he revolved continually over a nut fire. As the process continued a large ball or " biscuit " of rubber was built up on the stick, and then removed when it reached a weight of 40 or 50 pounds. It was in the

form of these biscuits that the wild rubber came upon the market. The coagulation process used on plantations today is a chemical one. The latex is put in large aluminum tanks which contain removable partitions. Dilute acetic acid is added. This has the effect of neutralizing the negative charges on the colloidal rubber particles and they

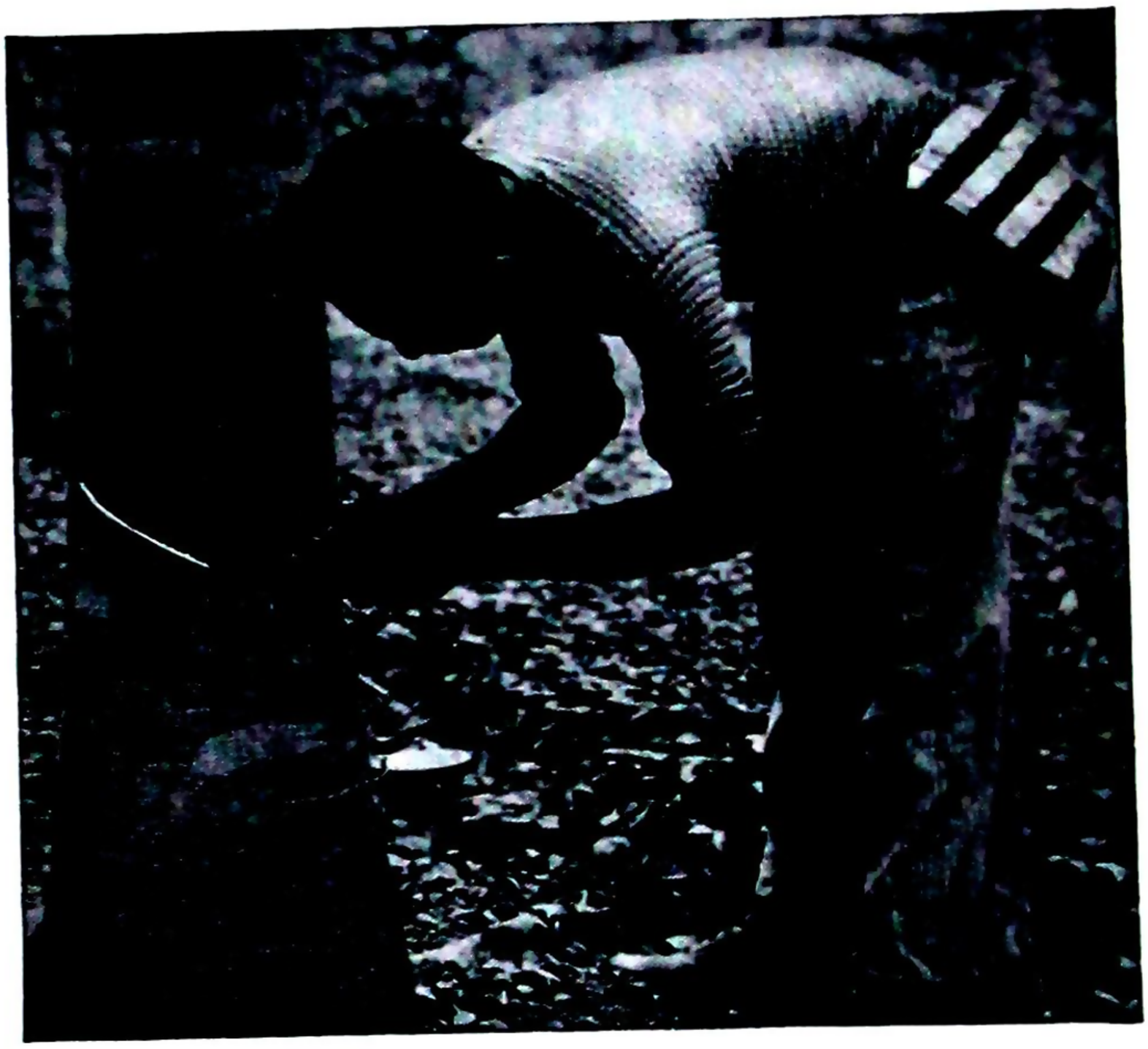


FIG. 98. Tapping a rubber tree. (Scientific American.)

collect together into a sort of semi-solid gel. When the reaction is complete the coagulated rubber or " coagulum " is removed from the tank in the form of a long, doughy strip.

The next step depends on which of the two most common types of crude rubber is desired. If it is to be of the sort called " pale crepe," the coagulum is passed between rollers somewhat as clothes are put through a wringer. This " wringer " differs from the ordinary one in three important particulars. First, the rollers are made of steel; second, there are a number of pairs instead of only one; and third, one member of a pair rotates faster than the other. The squeezing action, aided by a stream of water which plays on the rubber, serves to remove essentially all the impurities and liquids from the coagulum. The difference in roller speed produces a tearing action. The sheets

that emerge are hung in sheds to dry. They are yellowish-white in color and have the rough sort of surface that we associate with crepe rubber shoe soles.

The other kind of crude rubber is called " ribbed smoked sheet." In this case the coagulum passes between rollers which revolve at the same speed and serve simply to force out the liquids. Usually there is also some sort of design on the rollers which is impressed into the rubber sheets, hence the term " ribbed." Next, this rubber is hung in a smoke-house and given much the same sort of treatment as that accorded to hams and bacon on American farms. The finished product is dark brown in color and possesses an odor similar to that of the smoked meat mentioned above.

Until rather recently, almost all the rubber that left the plantations was in one or the other of these two forms. The last few years before the war, however, saw increasing amounts of liquid latex being shipped directly to the manufacturing plants of the United States. This development was made possible by the use of preservatives added to the latex to slow down bacterial action and to prevent the liquid from becoming acid, thus avoiding premature coagulation. The principal such preservative is ammonia. Before shipment the latex is concentrated so that the water content is reduced from 60 per cent to about 35 per cent. Important modern uses of rubber in this form will be described later in the chapter.

Addition of Compounding Ingredients. — Crude rubber must be given further treatment before it has any great industrial value. We have already made some reference to the nature and importance of vulcanization. In addition to the sulfur which is added for this reaction, various other so-called " compounding ingredients " are added to the crude rubber for a variety of reasons. The first step in this part of the process is breaking down the rubber to make it more plastic. This is done in a rubber mill, a device which consists principally of two heavy steel rollers arranged side-by-side. The rollers rotate toward each other with the back one turning more rapidly than the front one. The crude rubber is passed between these rollers time after time during both the " breakdown " process and the actual mixing in of the compounding materials. At the end of 10 or 15 minutes the former is complete. Then the compounding ingredients, consisting of oils,

tars, or other softening materials, as well as dry powders, are mixed in. This is accomplished by simply putting these materials on the rubber as it is being repeatedly worked through the mill. The kneading which the rubber receives serves to mix in the ingredients rapidly and uniformly. Now we will mention some of the groups of materials that are compounded with the rubber together with their functions.

Sulfur is added for the purpose of vulcanization. The actual vulcanizing process requires the application of heat and is usually not done until the finished product has been formed.

Accelerators are added to speed up vulcanization. The discovery of materials which would perform this service was immensely important because they have enabled the time necessary for vulcanizing to be reduced to a fraction of its former value. Some of them also permit vulcanization to be carried out at lower temperatures than would be possible otherwise. The most important accelerators are organic compounds.

Retarders are sometimes added in order to prevent vulcanization from taking place sooner than is desired.

Softeners tend to make the stock easier to plasticize and so reduce the time necessary for the milling operations. Among the typical softeners are vegetable oils, waxes, pine tar, and stearic acid.

Antioxidants, as the name implies, tend to retard oxidation and aging of vulcanized rubber caused by exposure to air and sunlight. They also hinder the deterioration which otherwise results from repeated stretching or subjection to heat.

Reinforcing agents, principal of which is carbon black, greatly increase the strength of rubber and the resistance it offers to wear. The fact that in 1941 automobile tires could be depended on to go from 25,000 to 40,000 miles whereas in 1910, 3,000 was considered a good mileage, is due to a considerable extent to the development and use of reinforcing agents.

Reclaimed rubber may or may not be added to the mixture, depending on the use to which it is to be put. So-called " reclaim " is produced by re-softening and re-plasticizing discarded rubber products. The material is not " de-vulcanized." Because reclaim costs less than a third as much as new crude rubber, its use reduces manufacturing costs. It should be emphasized that where reclaimed rubber

is employed, its use does not necessarily mean a reduction in quality. In fact for certain purposes the use of some reclaimed rubber adds desirable properties to the finished product.

Coloring agents are added to give the finished product the desired color. These include zinc oxide for white, lead chromate for yellow, and ferric oxide for red. Various organic dyes are also employed for this purpose.

Fillers of various kinds are added for such purposes as giving the mixture the proper consistency, producing a certain desired kind of surface, and lowering cost of manufacture.

Calendering, Frictioning, and Tubing. — The next operation depends on the form in which it is desired to have the final rubber stock. If sheets of rubber are wanted the compounded material is run through a machine called a " calender." This device consists of a series of smooth steel rolls whose distance apart can be accurately controlled. The same machine can be adjusted either to apply a thin coating of rubber on fabric, called " skim-coating," or to squeeze rubber into the fabric, known as " frictioning."

The tubing machine is one of the most useful of all rubber-working devices. Its operation resembles that of a sausage machine in that a worm arrangement turning in a cylinder forces the compounded rubber out through an opening or die. If a pin is placed in the center of the die-opening, the rubber will emerge in the form of a tube. This machine is used in manufacturing such articles as inner tubes, jar rings, garden hose, channel strips for automobile windows, and expansion joint fillers for concrete roads. It also is adapted for putting rubber insulation on wire.

Vulcanization. — As was mentioned above, this part of the process usually is not carried out until the compounded rubber has been fabricated into the form of the finished product. The exact curing procedure varies somewhat with the article involved. Tires receive what is called a " press cure." This means that they are placed in molds and heated while under pressure, the pressure serving to keep the tire from becoming porous because of enclosed gases. It is during this process that the tread design, which is on the inside of the mold, is impressed on the tire. Figure 99 shows a tractor tire being removed from the mold. Some products, such as fire and garden hose, insulated wire,

and thin sheet rubber, receive the heat necessary for vulcanization by being immersed in steam or hot water. Articles like raincoats and footwear are cured simply by being placed in an air- or carbon-dioxide-filled chamber which is maintained at a high temperature.

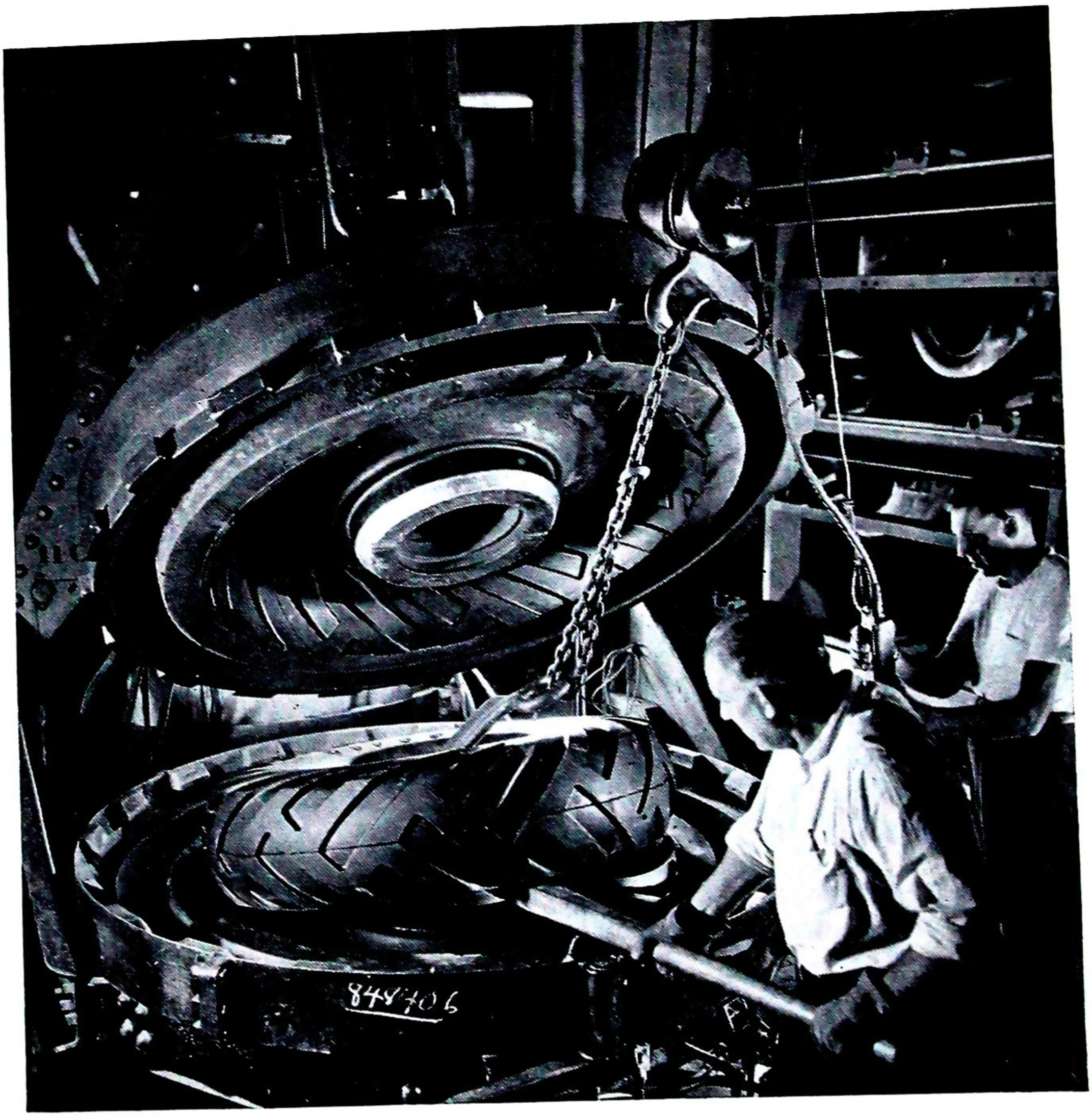


FIG. 99. View showing tractor tire being removed from mold. (Firestone Tire and Rubber Company.)

Temperatures necessary for vulcanization vary somewhat with the nature of the article but in general are between 230° F. and 290° F. The time that is required may be as little as a few minutes or as much as several hours, depending on what the product is and what accelerators have been used.

Electronic heating is now being used to vulcanize some types of rubber articles. This method of heating depends upon the heat developed when the rubber molecule is subjected to a high-frequency alternating current, which causes the molecule to bend rapidly, first in one direction and then in the other. During this process heat is

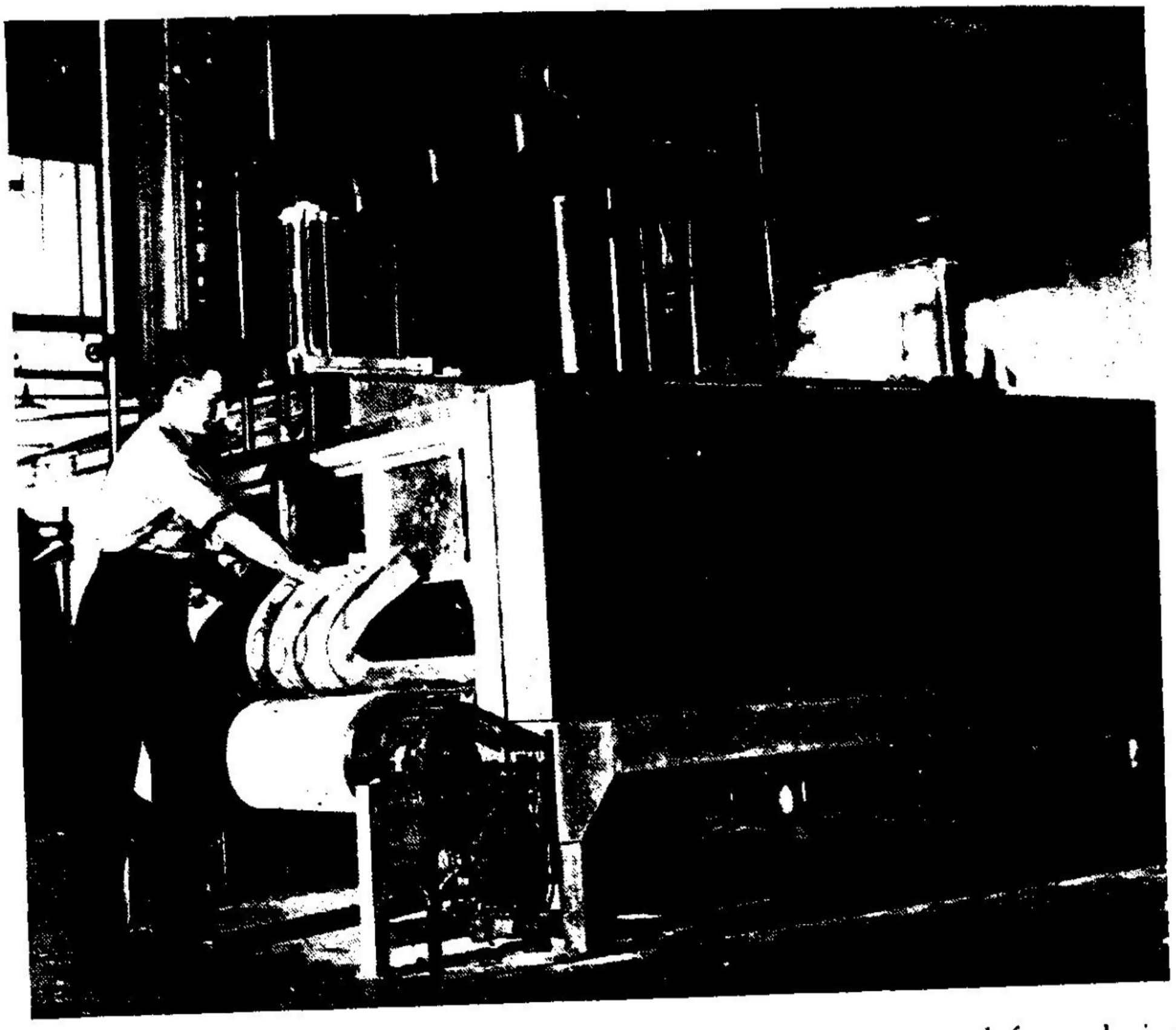


FIG. 100. Electronically-cured Foamex mattress being removed from drying oven. (Firestone Rubber and Latex Products Company.)

generated in much the same manner as occurs when a piece of wire is bent sharply back and forth. Electronic heating has not thus far been applied to vulcanization of tires but when it is the time of cure will very likely be reduced to a small fraction of its present value. One of the first applications was to the vulcanization and drying of rubber sponge. Figure 100 shows a foam rubber mattress being removed from an electronic vulcanizer and drier. This 125,000-watt unit, which was the first of its kind to be placed in commercial operation, is built on three levels including the vulcanization chambers, remote control panel and power unit.

Some articles are given what is known as a "cold cure." It has been found that if sulfur chloride is used instead of sulfur, the curing process will take place at ordinary temperatures. This process has found its principal applications in the manufacture of thin, bright-colored articles like balloons, bathing caps, and aprons.

Use of Cotton. — In addition to the rubber itself and the various compounding ingredients described above, cotton plays an important role in the rubber industry of today. Various kinds of cotton are employed in several different sorts of weaves, but in all cases it serves to give the finished product added body and wear-resistance. While we are most familiar with its use in the much publicized "cord tire," it also is found in garden hose, fire hose, footwear, belting, and many other articles. Its importance is evidenced by the fact that each year the United States tire industry uses around a third as many tons of cotton as it does of crude rubber.

Some Applications of Rubber

The fact that just before the war one rubber company made more than 30,000 rubber products makes it obvious that we cannot discuss in these pages all the applications of this material. The following list gives the distribution of natural rubber consumption in the United States for 1939:

	Per Cent
Tires	63.6
Tubes	9.2
Repair materials	3.1
Mechanical goods	9.0
Drug and sport sundries	5.0
Boots and shoes	6.8
Miscellaneous	3.3
	100.0

In our consideration of applications we will confine ourselves to a few of the more important and more recent developments in the field.

Tires and Tubes. — Before the war approximately three-fourths of all the crude rubber produced went into tires and tubes. The following figures indicate the extent of this production for 1940 in terms of units produced.

Tires	
Original equipment	22,261,723
Replacement sales	35,724,034
Export sales	1,169,569
Total	59,155,326
Tubes	52,350,867

Limitations of space prevent us from discussing here the particular procedures and techniques that are employed in the manufacture of automobile tires and tubes. Therefore, let us call the attention of the reader to the two references given at the bottom of this page.[1, 2] The story of tire manufacture is told and illustrated in these sources.

Direct Use of Latex. — Mention was made earlier in this chapter of the fact that increasingly large amounts of latex are being shipped directly to the manufacturer without first being made into pale crepe or smoked sheet at the plantation. There are many applications in which it is advantageous to work with the latex in liquid form. In such cases the compounding ingredients, including sulfur, are prepared in colloidal form and simply stirred into the latex.

One of the most important uses of rubber in this form today is in the manufacture of so-called latex sponge or foam rubber. In making this the compounded latex is put in a mixing device somewhat like a huge egg-beater, and whipped to a froth. It is then put in molds of the desired design and heated so that it vulcanizes. The material finds applications in mattresses, seat cushions, inner soles, athletic equipment, and many other products.

If the mixture, instead of being beaten up, is forced through a small opening into an acid bath, it coagulates to form rubber thread. Such thread has a circular cross-section and is greatly superior in strength and wearing qualities to the old square-cross-section thread that was made by slicing strips from vulcanized sheets.

The fact that the latex particles carry a negative charge is made use of in coating metal forms with rubber. One method of doing this, called " anode deposition," is very similar to electroplating, a process that will be discussed in a later chapter. Briefly, the procedure is to

1 Fisher, H. L., *Rubber and Its Uses.* Cleveland: Chemical Publishing Co., 1941.
2 Hildebrand, J. R., "Our Most Versatile Vegetable Product." *National Geographic Magazine*, February, 1940.

connect a source of electric current to two electrodes, the object to be coated with rubber being used as the positive one or anode. When the electrodes are placed in a container of latex and the current turned on, the negatively charged rubber particles are attracted to the positive form and deposited on it. Today, anode deposition has largely been replaced by dipping. The form to be coated is first covered with a solution of a salt. Then when the forms are dipped into latex, ions from the salt solution neutralize the rubber particles and they coagulate on the form. The forms that are coated may be irregularly shaped metal objects that are to be covered with rubber, or they may be glove or balloon forms from which the rubber is peeled off after a layer of the desired thickness has been built up. Products like rubber-coated canvas gloves and rubber-covered paper are also made by dipping.

New Uses for Rubber. — The following is a list of eight comparatively new applications of rubber which have been suggested by J. W. Schade, former Director of Research of the B. F. Goodrich Company, to illustrate some of the less familiar ways in which rubber is used as well as the great variety of its uses:

1. *Lining for Mills.* Reference has already been made to the great resistance which rubber offers to wear. This has made it an excellent material for lining mills used for such purposes as pulping paper or grinding up quartz ores. Additional advantages which it offers in this connection are decreased vibration and noise.

2. *Oil-well Casing Guards.* These consist of rubber collars which are placed around the drill pipe as it works in the casing to prevent wear of the latter. Here also the high abrasion-resistance of rubber is important.

3. *Rubber Bearings.* Rubber bearings have proved to be superior to metal ones in applications which involve water. In something like 85 per cent of the small boats made today, the propeller shaft turns in rubber bearings. In larger boats they are used in sizes up to 14 inches in diameter, while vertical bearings with diameters as large as 36 inches have been employed in water turbines. They outwear their metal predecessors many times, particularly in cases where sand particles work in along the shaft. An additional advantage lies in the fact that water is the ideal lubricant for rubber bearings.

4. *Rubber Putty.* A type of putty made from rubber adheres very tightly to both metal and glass and does not become brittle with age.

5. *Airplane De-icers.* Rubber has served to minimize the dangers to aircraft of ice forming on the wings, one of the greatest of all flying hazards. A rubber strip, placed along the leading edge of the wing, is so fixed that it can be alternately inflated and deflated. This serves to crack up any ice that may be collecting and permit it to be blown away.

6. *Tracks for Tractors.* Rubber-covered tracks for caterpillar tractors increase traction, decrease noise and at the same time give greater wear than all-metal ones. Many such tracks were used during the war on military vehicles.

7. *Upholstering Material.* The use of foam or sponge rubber in seat cushions and mattresses has already been mentioned. A somewhat similar application consists in combining rubber with animal hair to get an upholstering material. Such a product is highly resilient, light in weight, and gives long wear.

8. *Supports for Heavy Machinery.* This application resulted from the development within recent years of two methods of making rubber adhere tightly to steel. Rubber adheres well to brass of a particular composition, so one of the methods consists simply in first brass plating the steel and then sticking the rubber to the brass. In the other procedure, the steel is first painted with a particular type of cement, a rubber derivative, which serves to hold the two together. The accidental fashion in which this second method was discovered is of some interest. A group of investigators were searching for a rubber derivative from which phonograph records could be made. One compound that they tried stuck so tightly to the steel record mold that they were unable to get it free. Recognizing its possibilities in this connection, the experimenters promptly changed their interest to this material which could be used to make rubber adhere to steel.

The adhering of rubber to steel has proved particularly valuable in reducing shock and vibration caused by the operation of heavy machinery. Rubber pads placed under the machines have not been of any great advantage, so the present practice is to suspend the machinery from rubber. The procedure is to sandwich a layer of rubber be-

tween two steel plates, the materials being adhered together in the manner just described. The so-called " vibro-insulator " thus formed is then used in a vertical position, one steel plate being bolted to the machine and the other to the wall or other support. The machine ac-

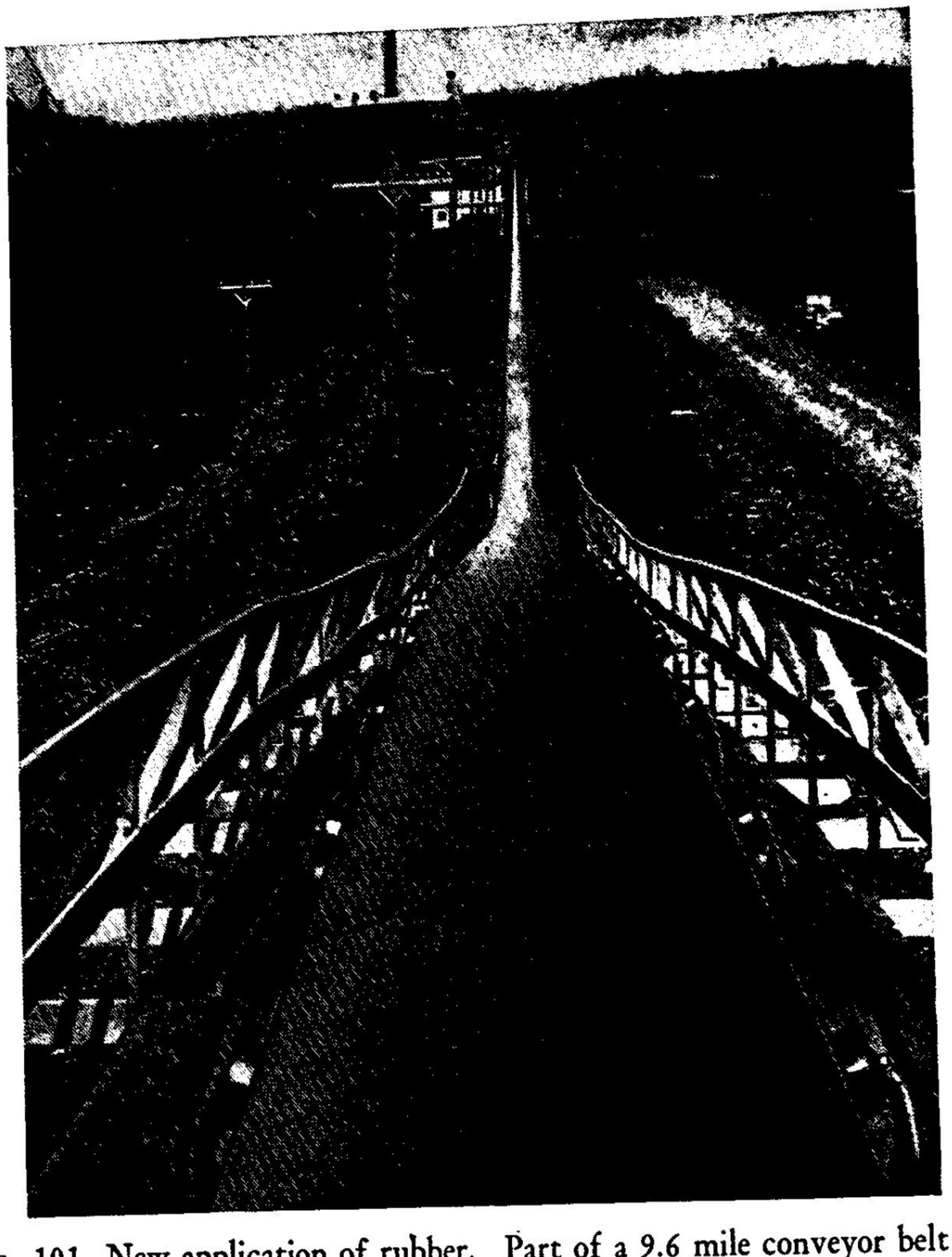


FIG. 101. New application of rubber. Part of a 9.6 mile conveyor belt system used to haul building materials for the Shasta Dam project in California. (Goodyear Tire and Rubber Co.)

tually hangs from rubber. By using a number of these insulators machines weighing as much as 80 tons have been suspended in this fashion.

Other applications of the bonding of rubber to metal include lining chemical tanks with rubber, and covering irregularly shaped objects with rubber.

Rubber Derivatives

Rubber derivatives are materials obtained by chemical reactions between rubber and substances other than sulfur. One of them is the cement already mentioned in the discussion of bonding rubber to steel. It is a product formed by rearranging the structure of the rubber molecule through chemical reactions which may be induced by a number of acidic compounds. Another material that is marketed as Pliofilm comes from the reaction of rubber and hydrogen chloride (HCl). It is less elastic than rubber and is readily formed into transparent sheeting or tape of a variety of colors. Outstanding among its desirable characteristics is the fact that it is highly resistant to oil and water. A third type of derivative, called chlorinated rubber, comes from reacting rubber and chlorine. From this there has been obtained a horn-like material which has proved valuable as a paint base.

A great deal of research has been done in recent years in this field of rubber derivatives, and it gives indications of future developments of enormous importance.

Synthetic Rubber

The term "Synthetic Rubber" is commonly applied to that group of materials having elastic and plastic properties similar to those of natural rubber. Since their composition, structure, and properties are not exactly the same as those of natural rubber, they are not truly synthetic rubbers. All rubbers, both natural and synthetic, are characterized by high molecular weight and are capable of being vulcanized.

Historical Background

The search for synthetic rubber began almost as soon as Charles Goodyear announced the discovery of vulcanization. Two groups of people became immensely interested in the problem of duplicating natural rubber. One group, for whom the field proved fertile, was composed of unscrupulous promoters who persuaded thousands of gullible investors to buy stock in companies that proved to be far more synthetic than their proposed product. The other group included the chemists who recognized the real value that synthetic rubber could have, and set about trying to solve the problem.

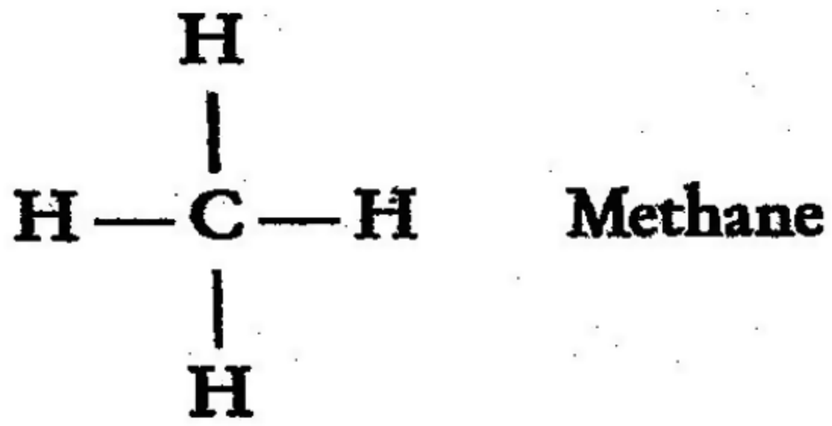
Obviously, to synthesize rubber, it was necessary to know accurately the composition of the natural material. Michael Faraday, an English scientist, had established in 1826 the ratio of carbon to hydrogen atoms in the rubber molecule and had given it the formula C_5H_8 which, as we have seen, is now written as $(C_5H_8)_x$. In 1860 another Englishman, by heating rubber, broke it down to form a colorless, low-boiling liquid, the composition of which was also found to correspond to the formula C_5H_8. This he called isoprene. Then in 1879 a French chemist converted isoprene back into an elastic mass approximately like rubber. But it was not identical with it, and so synthetic rubber had not quite been produced.

The attempts to duplicate rubber by starting with isoprene were continued for many years, particularly in England and Germany. Although in both countries products were obtained which had some extensibility, they were all of very poor quality with respect to aging and wear resistance. No important commercial use was made of these substances until Germany's supply of natural rubber ran low in 1915 because of the English blockade. At that time a German factory was put into operation producing synthetic rubber, but the product was still very inferior and found almost its only use as a substance from which to make articles of hard rubber, such as battery boxes for submarines.

Following the war, the search was continued but still without appreciable success, as far as finding a satisfactory substitute for natural rubber was concerned. In fact, no progress was made until chemists stopped trying to duplicate the rubber molecule and turned their attention to materials that would have similar physical properties even though chemically they might differ from it.

In order to be able to point out the degree of resemblance between the synthetic and the natural substance, it is necessary that we introduce the reader to a somewhat different kind of chemical formula than that employed so far in this book. Consider, for example, methane gas, a compound that in an earlier chapter we designated as CH_4. Now this formula tells only what atoms are present in the molecule. The organic chemist finds it desirable to use a type of formula that represents not only what constituents are present, but also how

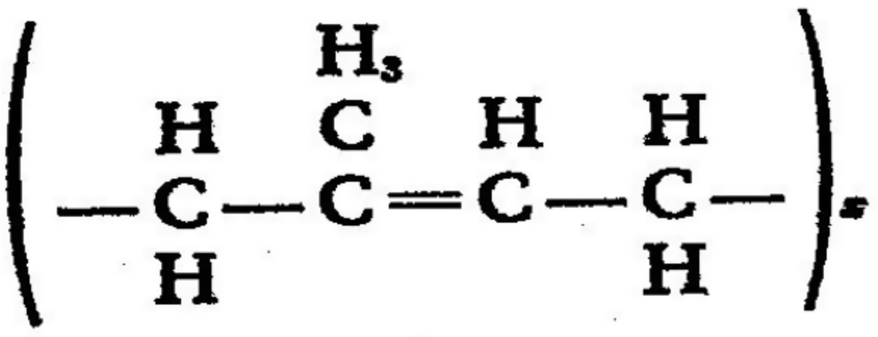
they are arranged. So he writes for methane what he calls a graphic or *structural* formula, as follows:



Methane

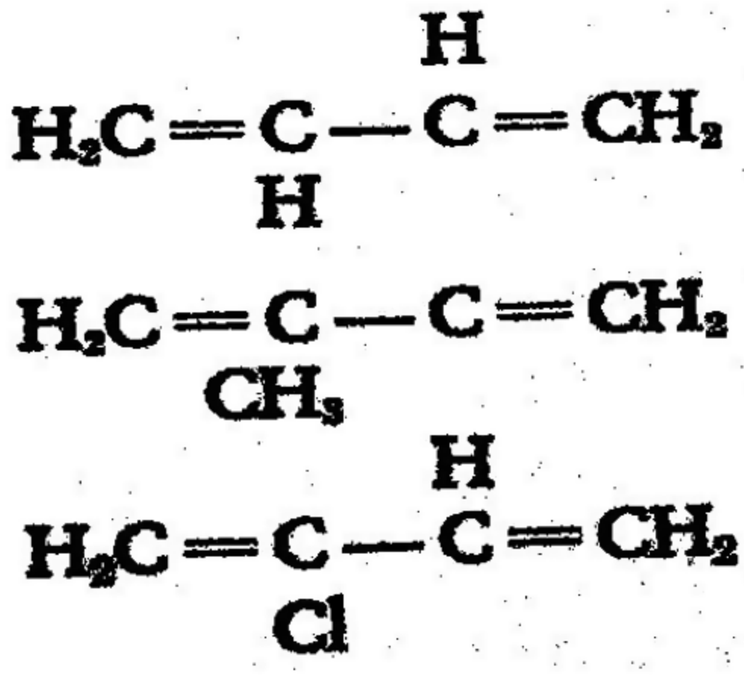
where the short straight lines represent bonds or imaginary "hooks" which can be thought of as indicating points of attachment of other atoms to carbon. The number of these bonds is the chemical valence of the element.

The formula for the rubber molecule, which we have written so far as $(C_5H_8)_x$, in its structural form is expressed as,



Rubber

where the unattached bonds at the ends are used to attach similar molecules to form the large molecule which is called the "polymer." Isoprene, which has the formula C_5H_8, forms big molecules composed of the same units as rubber. From the structural formulas of butadiene, chloroprene, and isoprene, given below, their similarities and differences can be noted.



Butadiene

Isoprene

Chloroprene

It should be noted that in each of the above molecules there are two double bonds, while in the unit of natural rubber there is only one and it is in a different position. This arrangement of the double bonds occurs with butadiene and chloroprene the same as with iso-

prene and is the fundamental mechanism by which the small molecules add on to each other and form large ones, that is, "polymerize."

If all the small molecules are alike they are said to form polymers; if they are different, as with butadiene and isoprene, they form copolymers and the process is called copolymerization.

It will be shown in the next chapter that this same mechanism is involved in the manufacture of synthetic plastics. However, the manufacture of big molecules from little ones can be accomplished by other reactions which are not involved in the manufacture of synthetic rubbers.

When the Rubber Reserve Corporation was assigned the task of providing the United States with a supply of rubber independent of foreign sources, it had to answer several important questions, namely,

1. What synthetic rubber was the best substitute for natural rubber in its most important uses, e.g., tires?
2. Was there a sufficient quantity of raw materials available to manufacture this type of synthetic in the necessary amounts?
3. Could this type of synthetic be manufactured by available equipment and did the United States possess the necessary "know how"?
4. What special purpose synthetics were needed and how would their manufacture fit into the over-all picture?

It is not possible here to go into the reasons behind the decisions that were made but the types finally decided upon were the following:

1. A butadiene-styrene copolymer (GR-S)[3] was selected as the general purpose rubber.
2. A butadiene-acrylonitrile copolymer (GR-A) was chosen as the oil resistant rubber.
3. Polychloroprene, or neoprene, (GR-M) was also selected for both oil and sunlight resistance as well as a substitute for general purposes.

[3] GR-S—Government Rubber—Styrene (copolymer).
GR-A—Government Rubber—Acrylonitrile (copolymer).
GR-M—Government Rubber—Monovinyl (acetylene polymer).
GR-I—Government Rubber—Isobutylene (isoprene copolymer).

4. Butyl rubber (GR-I) was considered experimental at first but was later incorporated into the program. It is a copolymer of butylene and isoprene.

All of the raw materials are obtainable from petroleum, coal, salt, limestone or combinations of these. It was soon discovered that the manufacture of butadiene, which was needed in the largest quantities, would interfere with the production of aviation gasoline if it were to be prepared from petroleum. It then became a question of making butadiene from petroleum and doing without gasoline or making it from alcohol and doing without whiskey. The program was switched to making a large part of the needed butadiene from alcohol. Neoprene is included in this chapter as a synthetic rubber, in spite of the fact that it is not vulcanized with sulfur like the others, because its physical properties are very similar to those of natural rubber. It was developed by duPont chemists from discoveries of Father Julius Nieuwland of the Chemistry Department of Notre Dame University.

Properties and Applications

The compounding of the butadiene elastomers is essentially the same as for natural rubber except for the amounts of the ingredients required. Because, in the case of butadiene-styrene copolymer GR-S, much larger quantities of carbon black are needed than for the natural rubber, a serious shortage of that ingredient developed. In the case of neoprene, certain metallic oxides, when mixed with the rubber, produce on heating a change similar to that in natural rubber when it is heated with sulfur.

It has been estimated that synthetic rubber could replace the natural product in 95 per cent of the latter's uses. Although no synthetic rubber is as versatile as natural rubber, there are enough different kinds, each with its own set of properties, that one can be found to meet almost any set of specifications. The new synthetic rubber tires have been developed to a point such that some manufacturers claim they will wear better than tires of natural rubber. Inner tubes made of butyl rubber are much better than natural rubber in their ability to hold air. Self-sealing gas tanks for combat aircraft are made by putting a layer of oil-soluble synthetic rubber between

two layers of an oil-resistant rubber of the butadiene-acrylonitrile type. When a bullet penetrates the tank, the gasoline comes into contact with the oil soluble rubber which immediately begins to swell and close the hole. This same oil-resistant synthetic is used for gasoline and oil hose as well as for oil pump gaskets and diaphragms. Neoprene, which is particularly valuable where sunlight resistance is an important factor, replaces the natural rubber in other applications also.



FIG. 102. An automobile tire made of Ameripol. (The B. F. Goodrich Company.)

It was announced in January, 1946, that natural rubber, when it was again available, would cost about 25 cents per pound. At the peak of synthetic rubber production, costs in some plants were claimed to be as low as 11 cents to 12 cents per pound for the GR-S rubber, 19 cents for neoprene, 13 cents to 18 cents for butyl, and 18 cents for the acrylonitrile type. Costs in peacetime will depend to a large extent on the costs of raw materials on an open market. Dr. Egloff of the Universal Oil Products Company has estimated that butadiene can be made in this country to provide an annual production of 85,000,000,000 pounds of synthetic rubber, while recent estimates are that the total world consumption of natural and synthetic rubber in postwar years will be in the neighborhood of 3,000,000,000 pounds. This fact together with the expansion of rubber plantations in this

hemisphere would seem to indicate that the United States is rather favorably situated to become independent of a rubber supply that is thousands of miles from its own shores.

Questions

Discussion

1. Describe a method for determining whether or not any given liquid is a colloid.

2. How many microns in a centimeter? in two inches?

3. As nearly as possible, find an example other than the one given in this chapter for each type of colloid listed.

4. Name and discuss four properties of colloids.

5. The rubber that may now be part of a tire was in a colloidal state at what stages in its history?

6. Look up information about the use of rubber before the time of Charles Goodyear.

7. Look up information about the first rubber company to be started in this country.

8. Mention reasons for Hevea braziliensis being considered the best crude rubber source.

9. Describe the early development of rubber plantations.

10. Why is the chemical formula for rubber written $(C_5H_8)_x$ rather than simply C_5H_8?

11. What are the constituents involved in the reaction called vulcanization?

12. What physical properties of rubber make it so valuable a material?

13. What is the principal undesirable property of rubber?

14. Describe the various plantation operations in the processing of rubber.

15. Identify the following terms as they apply to the preparation and processing of rubber: coagulum; pale crepe; smoked sheet; rubber mill; calender; anode deposition.

16. For what purposes are the following compounding ingredients added to crude rubber: sulfur; antioxidants; retarders; accelerators; carbon black?

17. Mention and discuss four of the so-called new uses for rubber given in this chapter.

18. Find as many other recent applications of rubber as you can.

19. Distinguish between polymerization and condensation.

20. Find from other sources the structural formulas for several organic compounds not mentioned in this chapter.

21. Name several sources of butadiene.

Multiple Choice

1. A suspension of a solid in a liquid, in which the diameters of the solid particles were around two or three microns, would be classed as (a) a colloidal suspension, (b) a cloudy suspension, (c) a coarse suspension, (d) a suspension of massive particles.

2. Which one of the following colloidal properties is made use of in certain smoke elimination devices? (a) Colloids exhibit Brownian movements. (b) Colloids exhibit the Tyndall effect. (c) Colloids are stable. (d) Colloidal particles are electrically charged.

3. The name of Charles Goodyear is associated with the (a) development of the first rubber plantation, (b) smuggling of Hevea braziliensis out of Brazil, (c) invention of the rubber mill, (d) discovery of accelerators, (e) discovery of vulcanization.

4. Which one of the following is out of place? (a) Coagulum. (b) Accelerator. (c) Reinforcing agent. (d) Activator. (e) Antioxidant.

5. Which one of the following is involved principally in vulcanization? (a) Carbon. (b) Cotton. (c) Zinc oxide. (d) Stearic acid. (e) Sulfur.

True-False

1. All colloids exhibit Brownian movements.

2. The name of Tyndall is associated with a device used for the elimination of smoke.

3. The molecular weight of rubber is believed to be many times greater than the number that would represent the sum of the atomic weights of five carbon and eight hydrogen atoms.

4. The fact that the particles in a given colloid all carry the same kind of electric charge helps explain why they do not settle out.

5. Coagulation of rubber is usually accomplished by electrolysis.

6. Before reclaimed rubber is used all sulfur is removed from it, that is, it is " de-vulcanized."

7. One of the processes of bonding rubber to steel involves first plating the steel with brass.

8. Rubber belongs to that group of substances which do not have sharply defined melting points.

9. Koroseal is a rubber derivative, that is, is made by reacting rubber with certain chemicals.

10. Most synthetic rubbers offer greater resistance to oil than does natural rubber.

Chapter 19

SYNTHETICS

THE dictionary defines " synthesis " as " the putting of two or more things together." By this definition a " synthetic " would be the product that results from the " putting-together " process. The great variety of materials that we group under this classification today is properly described by the above definition if certain qualifications are placed on the process. For instance, the putting together of the constituents that go into modern synthetics is never just a simple physical mixing-up. It is, generally, a rather highly complicated procedure involving a number of steps and various chemical changes. There are two general reasons why man has tried to make these materials instead of being satisfied with the ones provided by nature. One is his desire to duplicate natural substances either to reduce cost, or because of a shortage of the natural product, or both. The other reason is an attempt to produce materials with combinations of desirable properties not found in the things provided by nature. Very frequently the second objective has been obtained as a result of experiments begun for the first reason.

Since, for purposes of this discussion, we are going to divide the subject into inorganic and organic synthetics, it is desirable that we define these terms. The distinction between organic compounds and inorganic compounds may be made very simply by saying that the former contain carbon and the latter do not. This, of course, immediately raises the question as to why carbon, instead of oxygen, or hydrogen, or some other element, is made the criterion. The original reason was the belief that carbon-containing compounds could only be formed as a part of " life processes," an idea that has since been shown to be incorrect. Of the various valid reasons that can be given for the

importance of carbon, the one that particularly concerns us in this connection is the ability that carbon atoms have to combine with each other to form long chains. These chains, to which atoms of other elements may be attached, then form the skeletons, as it were, of organic compounds. While carbon is not unique in this respect, it does possess the property to a greater extent than does any other element, a fact which makes possible the great number of known organic compounds, now numbering half a million or more.

Inorganic Synthetics

We will consider individually only two of the most important inorganic synthetic materials, glass and ammonia. Others include sulfuric acid, Portland cement, and synthetic sapphires, rubies, emeralds, amethysts, and topaz all having the same chemical composition as the natural gems.

Glass

Glass is undoubtedly the oldest man-made synthetic. It can be called a synthetic plastic because its internal structure is quite similar to that of the organic materials usually included under that term. No one knows just when or under what conditions glass was first made. The process may have been discovered by Egyptian priests, although Pliny credits it to some Phoenician sailors. The earliest actual evidence is in the form of objects having about the same composition as ordinary bottle glass, found in Egyptian tombs dating back to 3500 B.C. In spite of the industry's great age, scientific research into ways and means of improving the product has only been carried out within the past half century or so, and many of the most important developments have occurred within the last 20 years.

Plate glass, and glass for windows and bottles constitute the major portion of the industry's production. The composition of these three is about the same, being principally the oxides of silicon, calcium, and sodium, which are obtained from the raw materials, white sand, limestone, and soda ash respectively. Silicon oxide (SiO_2) makes up about 70 per cent of the average batch, calcium oxide (CaO) some 13 per cent, and sodium oxide (Na_2O) most of the remainder. The ingredients are fused together at a temperature of about 2,700° F.,

then cooled to around 2,100° F., and fabricated into the desired product.

Low-expansion glass represents one of the most important modern developments in this field. The material that we know as Pyrex expands only about one-third as much as ordinary glass for a given temperature change. This property makes possible its use in cooking utensils or for other applications where it is subjected to comparatively rapid temperature variations. More recently there has been developed a glass that is so resistant to sudden temperature changes that it will stand having molten lead poured on one side of it while the other side is in contact with ice. This glass is made by dissolving out with acid a part of the components of a normal glass. Since the silica is insoluble in acid it remains behind. The remaining glass is then heated until it fuses together to form a substance having a very high silica content and properties of heat resistance and light transmission almost equal to those of pure quartz.

Some developments in the glass industry involve the substitution of other ingredients for silica. One product, in which phosphorus pentoxide replaces silica, unlike ordinary glass, is not affected by hydrofluoric acid or fluorine. When beryllium fluoride is used in place of silica the resulting glass transmits light more rapidly than does water. Glass building blocks, glass fibers and glass cloth are among the other glass developments which show great promise. The last of these is finding important applications in the electrical and heat insulating fields as well as in the production of plastic laminates.

Synthetic Ammonia

Ammonia (NH_3) was first isolated and studied as a chemical compound by Joseph Priestley, the discoverer of oxygen, carbon dioxide, and carbon monoxide. This occurred in 1774. The importance of ammonia as a plant food was not recognized until 1850, and since then increasingly large quantities of it have been used in commercial fertilizers. It also finds important applications in the manufacture of explosives.

Before the development of methods for making ammonia synthetically, two sources accounted for almost all of the world's production. Until shortly before the beginning of the present century most

of it was made as a by-product of the manufacture of coke. Since about 1895 some ammonia has been manufactured by what is called the "cyanamid process." This is an efficient method, requiring cheap electric power and using as raw materials limestone, coal, nitrogen, and water.

Direct synthesis of nitrogen and hydrogen to form ammonia was first made possible commercially by the German chemist, Fritz Haber, in 1913. Without this development, the 1914–18 World War probably either would never have started or would have been over within a few months after the English blockade cut off Germany's supply of Chilean nitrates.

As the method is used today in the United States, a mixture of nitrogen and hydrogen is passed through a catalyst of various metallic oxides at a temperature of some 900° F. and under a pressure of 5,000 pounds per square inch. Under these conditions synthesis is accomplished.

In 1938 the world production of ammonia was something over 3,000,000 tons, of which five-sixths were made synthetically. The estimated production in the United States during 1945 was between 500,000 and 600,000 tons. It has been estimated that about 200,000 tons of ammonia would be required annually to meet the explosive needs of a 2,000,000-man army.

Organic Synthetics

In the preceding chapter we have discussed those synthetics which are similar to natural rubber in that they are of high molecular weight, elastic, and capable of being vulcanized. In this chapter we shall consider other synthetics which also have high molecular weights but which cannot be vulcanized and which possess little if any extensibility unless modified by plasticizers.

General Characteristics of High Molecular Weight Compounds

By high molecular weight compounds we mean those whose molecular weights are in the range from 10,000 to 1,000,000. Physically they may vary from hard, almost crystalline, solids to rubbery, sticky, and soft materials. Each molecule is built up of a number of basic units which are repeated many times to make up the giant molecule.

We still cannot say exactly why so great a variation exists among the properties of the different kinds of large molecules, but our knowledge is growing constantly. The following analogy has been suggested as an aid in visualizing the internal structure of these molecules as we think of them today. Everyone has seen a brush pile made up of the branches of various sizes and shapes which accumulate as a tree is trimmed. If one jumps on such a heap of branches he discovers that it exhibits a considerable amount of resilience or " bounce." A single branch can be pulled from the heap but it tends to bring others with it and will spring back if released. In this state the brush pile is analogous to unvulcanized rubber. Next we will assume the branches to be tied tightly together at all their points of contact in the pile so that they cannot be pulled apart although the heap is still resilient. In this state the brush pile is analogous to vulcanized rubber. If we substitute rigid steel rods, not tied together, for the flexible wooden branches, we have the analogue of a hard plastic like an insoluble infusible resin.

The common distinction between elastics and plastics on the basis of their relative properties of extensibility and hardness is largely an arbitrary one which depends on their behavior within the normal temperature range. For example, if a piece of rubber is kept at 0° F. for some time it becomes stiff and leathery; if it is cooled to the temperature of dry ice or liquid air, it becomes hard and brittle like glass. If polystyrene or polyvinyl chloride, which we think of as hard plastics, are heated to the temperature of boiling water they become soft and rubbery. Much of the development work with elastomers and plastics has been for the purpose of extending their range or usefulness to higher and lower temperatures. This was especially true during the war when rubber on a single airplane would be subjected to the heat of the desert one day and subzero temperatures over the " Hump " the next.

Synthetic Plastics

Man has known how to make all sorts of *things* for many centuries. Until about 1870, however, for materials out of which to make these things, he was limited to nature's products, such as wood,

stone, metals, horn, tusks, bristles, natural resins, and vegetable and animal oils. The only exception to this was glass, which we have already mentioned as being the first plastic. In using these available raw materials, man has had to resign himself to the particular combination of properties .that they happened to possess, regardless of whether or not they were desirable characteristics. With the development of modern plastics, however, he has learned how to make many raw materials out of which other things can be constructed and, within limits, is able to put into these materials the properties he wishes them to have.

Definition of a Plastic. — One author suggests that there are three conditions a material must fulfill if it is to be classed as a modern synthetic plastic. First, it must be a synthetic substance that is built up chemically and is not found in nature. (To paraphrase a familiar quotation, only man can make a plastic.) Second, the molecules of the substance must be able to polymerize. And third, at some stage of the manufacture, it must be capable of flow under heat, pressure, or both.

General Considerations. — As the above definition suggests, plastics did not just happen. Their present development is the result of what probably has been the most extensively planned research program in the history of American chemistry. This research has been pointed toward the production of the following five main types of products:

1. Synthetic resins for use as plastics in the manufacture of molded articles.
2. Synthetic transparent resins.
3. Synthetic bases for paints and lacquers.
4. Synthetic resins for adhesives.
5. Synthetic fibers.

There are a number of bases upon which plastics may be classified including chemical composition, chemical reactions by which formed, the forms in which they are supplied to the public, the physical properties of the finished products, and raw material sources. We shall use as our basis the chemical reactions by which the materials

are formed and at the same time we shall attempt to show their relation to each other with respect to raw materials.

Chemical Reactions Involved in the Manufacture of Plastics. — There are two general types of chemical reactions by which plastics are formed. The first of these is the usual one of polymerization by which small molecules unite with each other to form large molecules. This is the process, described in the preceding chapter, by which synthetic rubbers are made. The most important plastics formed by polymerization are those of the vinyl type, all of which may be said to be based on acetylene, although acetylene is not necessarily an important link in their manufacture. The table below shows the chemical formulas of the basic units of the different vinyl polymers as well as the scientific name, a commercial name, and some typical applications.

TABLE 20. CHEMICAL FORMULAE AND NAMES OF VINYL POLYMERS

Formula	*Scientific Name*	*Commercial Name*	*Typical Applications*
H H —C—C— H H	Polyethylene	Polythene	Insulation
H H —C—C— H Cl	Polyvinyl chloride	Koroseal	Molded goods, insulation, waterproof cloth.
H Cl —C—C— H Cl	Polyvinylidene chloride	Saran	Bus seat covers, molded goods, tubing.
H H —C—C— H Ac	Polyvinyl acetate	Vinylite	Insulation, molded goods, waterproof cloth.
H H —C—C— H 	Polystyrene	Lustron (transparent)	Electric equipment, insulation, molded goods.

Another type of vinyl resin is the polyvinyl acetals which are used extensively in automobile windshields. The basic unit is much more complicated than those listed above and for that reason is not given in the table.

The vinyl resins are thermoplastic which means that they will flow under pressure at about the temperature of boiling water. At ordinary temperatures they are hard plastics although they can be modified by the addition of plasticizers to form soft rubberlike mate-

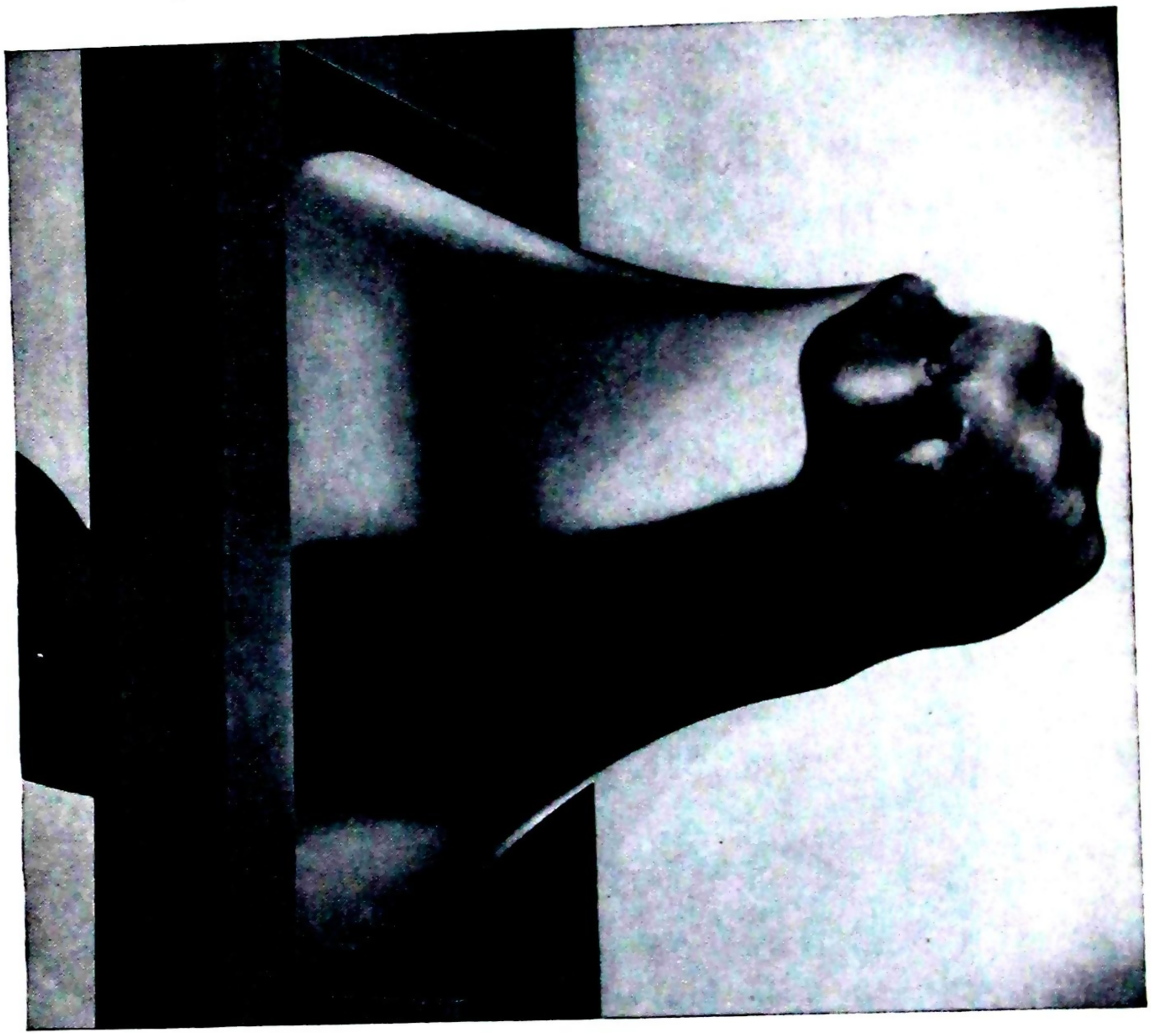


FIG. 103. Photograph illustrating the " stretchability " and toughness of the new vinyl resin plastic now being used in making safety glass. (Libbey-Owens Ford Glass Co.)

rials which can be substituted for rubber in a great many uses. Other applications of the vinyl resins are in molded goods, extruded articles like wrist watch straps, belts, tubing, and electric wiring insulation. Naval ships and aircraft are wired with electric equipment insulated with polyvinyl chloride and polyvinyl acetate. They can be dissolved in solvents and spread on cloth to make it waterproof for use in shower curtains and raincoats. They can be calendered into sheets

and sealed to make waterproof containers or pressed into paper to make oilproof containers to replace tin cans.

Closely related to the vinyl plastics are the acryloid plastics which are so named because they are derived from acrylic acid. These form the most transparent plastics and include in their applications bomber noses and turrets and fighter plane covers. They will undoubtedly replace glass in many uses as, for example, the contact type of eye glasses in which the lens fits against the eyeball. Many molded articles

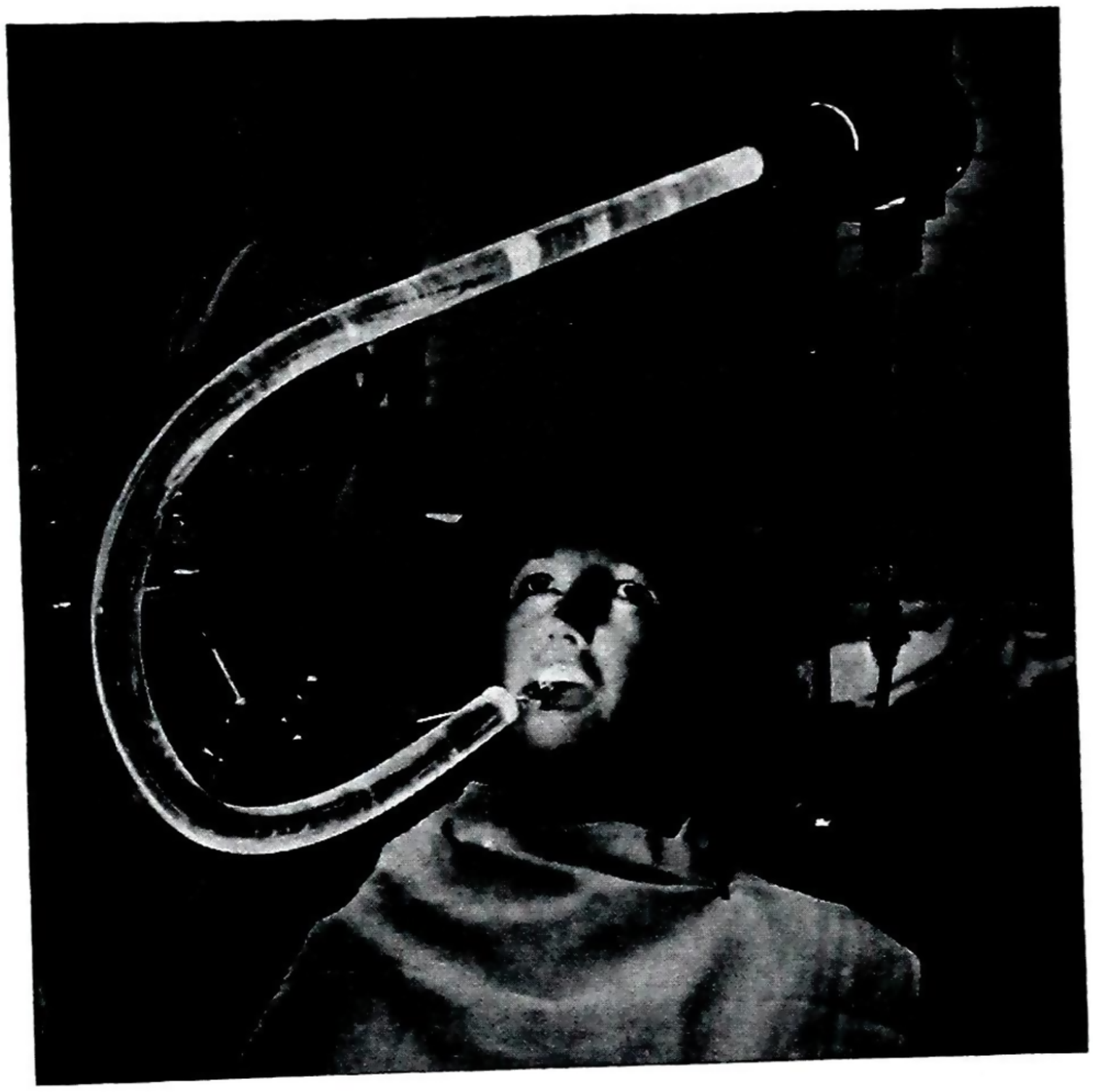


FIG. 104. Photograph illustrating one application of the " piped illumination " that is possible with Lucite. (E. I. duPont de Nemours and Co., Inc.)

are made from transparent plastics of this type, including furniture, picture frames, dresser sets, and others.

One of the most interesting properties possessed by acryloids is one which makes possible the phenomenon known as edge-lighting, illustrated in Fig. 104. As this photograph indicates, light which enters one edge or end of a bar of Lucite travels through it regardless of the turns encountered and emerges almost undiminished from the other end. That is, light is carried through a bar made of this material in a fashion somewhat analogous to the way water is carried

through a pipe. Applications include advertising signs, illuminated dental and surgical instruments, and illuminated batons and pointers.

The second general type of reaction is that in which the linking together of a large number of basic units is accomplished by a reaction

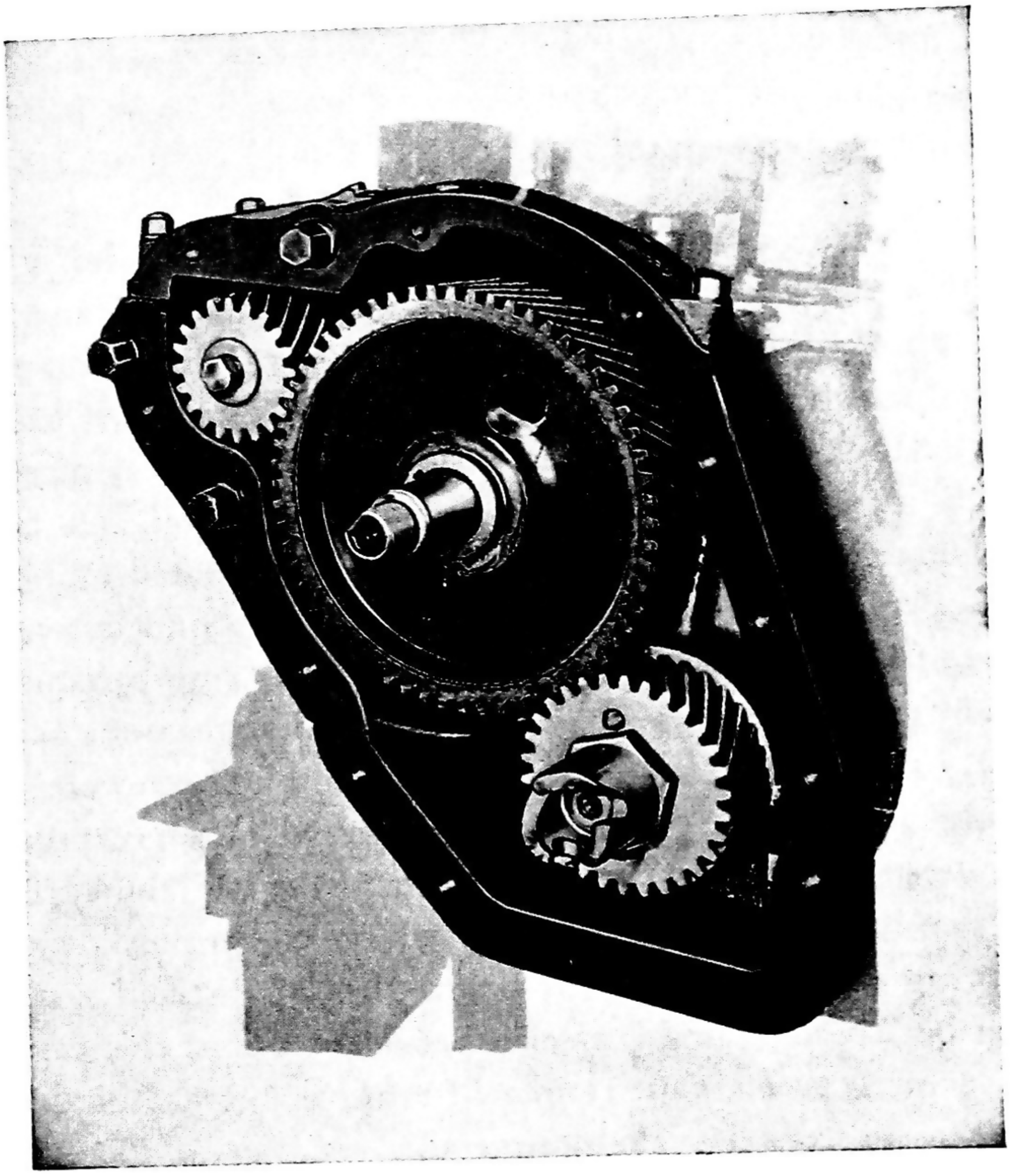


FIG. 105. A laminated automotive timing gear between steel gears. (Bakelite Corporation.)

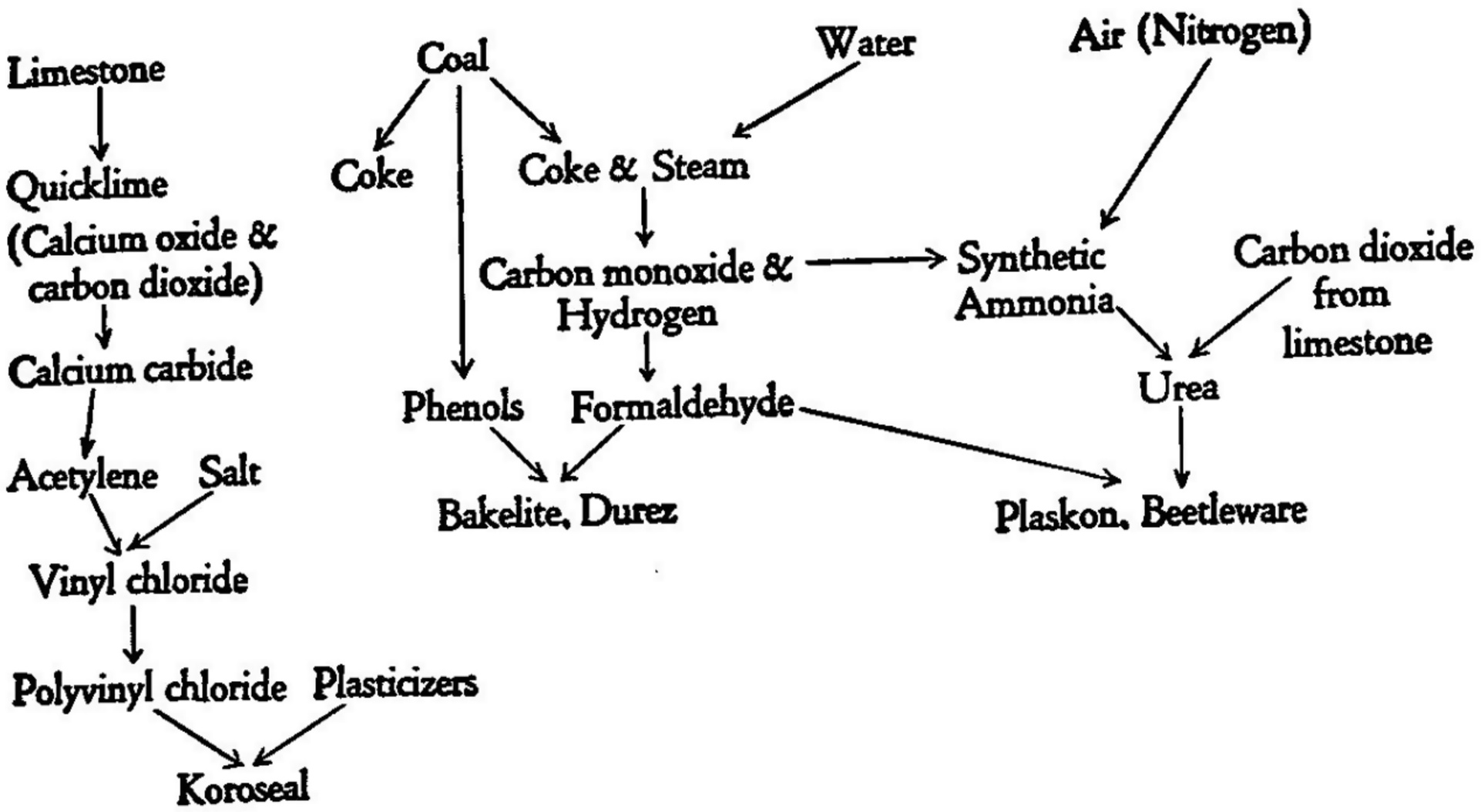
between two different molecules with an accompanying splitting off of another molecule such as water. These form the thermosetting resins, so-called because when heated they harden into infusible, insoluble resins. They first form an intermediate stage in which they are thermoplastic or in which they may be mixed with other materials to give special properties to the finished product. They may be molded into any shape and then cured into the hard form. It is this

two-stage process that makes them useful as adhesives and bonding agents. One member of this type is Thiokol in the preparation of which table salt instead of water is split off. The three widely used groups of resins which are formed by the kind of reaction just described are,

1. Phenol-aldehyde resins.
2. Urea-aldehyde resins.
3. Alcohol-acid or alkyd resins.

Resins of this type find application in molded articles, as Bakelite and Beetleware, as lacquers and paints, and as adhesives and bonding agents. They are used to bond together paper, fabric, and plywood. Paper laminates can be molded to form gears (Fig. 105), tubing and other articles as well as electric insulation. Fabric laminates have been used in airplane construction to replace metal. A core of a plastic sponge is placed between two layers of laminated fabric bonded together with a thermosetting resin. Plywood bonded together by this type of resin has many uses. If completely impregnated with it the wood assumes a highly polished, varnished appearance and can be used for furniture. The famous Lancaster bombers were made of such bonded plywood. Other types are used in a great variety of wood construction. Some of these resins have the ability to absorb part of the mineral content of water which makes possible a method of purifying water at a cost far below that of distilled water. The resin can be " regenerated " after a period of use in this fashion and made as good as new. One resin will remove the metallic substances in the water and another the nonmetallic.

Raw Materials. — Since the plastics industry today produces and uses millions of pounds of these synthetics annually, it is important to know something about the raw materials and how they are made into resins or plastics. The following chart shows what raw materials are used in making representative members of each group of plastics discussed and the important steps in their manufacture.



Synthetic Fibers

Ever since man has covered himself with cloth instead of the skins of animals he has had three basic fibers from which to choose, namely, cellulose, wool, and silk. The cellulose fibers were linen and cotton. The wool came chiefly from sheep, though some was obtained from goats and llamas. Silk came from the silkworm and was for centuries worn almost exclusively by royalty. It was only natural that man should attempt to make a silk that would displace the costly natural variety and make it possible for the poor man to clothe himself in this kingly material. Among the first such experiments were those directed toward making silk from proteins. The story is told of a professor of organic chemistry who set out to prove that silk purses can be made from sows' ears and was able to make silk fibers from proteins and have them fashioned into a silk purse. Recently fibers have been made from proteins in milk and it probably is only a matter of time until the vegetable proteins will be used. The first successful "artificial" silk was made from cellulose. Today viscose rayon is the most important such fiber. Its chief disadvantages are its loss of strength when wet and its slow drying properties. Also, it lacks certain other desirable properties of real silk. Cellulose acetate is made into a fiber sold under the trade name of Celanese.

The first real synthetic fiber was Nylon. This was an outgrowth of the work of Dr. Carrothers, a duPont chemist who did some of the earliest and most fundamental work on polymers in this country. Nylon is a polymide and is formed by a type of reaction somewhat different from those discussed earlier. It forms a fiber which has considerable stretch and recovery. It is the strongest fiber known and was used extensively in the recent war for parachutes and for tow ropes for gliders. Loaded gliders could be taken in tow by planes while in flight without the tow line breaking and with minimum shock to those in the glider. Nylon fabric has been used in tires and in other rubber goods although some difficulty has been experienced in making it adhere successfully to rubber. A new Nylon has recently been reported to have an extensibility of as much as 300 per cent. The chief advantages of Nylon are its durability and toughness and its resistance to water and to high temperatures. It can be boiled and sterilized without being damaged. It is in great demand for women's hosiery because of its resistance to "runs." Research is in progress which no doubt will produce other fibers as good or better than Nylon.

Other Synthetics from Vegetable and Animal Sources

Synthetics other than fibers are derived from vegetable and animal materials. The plastic which we know as Celluloid, or cellulose nitrate, which was developed about 1886 to replace ivory in the manufacture of billiard balls, is made by treating cellulose with nitric acid. For many years it was used extensively for all sorts of molded articles including toys, dresser sets and collars. Photographic film also was made of cellulose nitrate. Its very high inflammability, however, makes it very dangerous. Today it has largely been replaced by cellulose acetate which is made by treating cellulose with acetic acid. This material closely resembles the nitrate in appearance, but is slow-burning. It is used for photographic film, a great variety of molded articles and some transparent plastic materials. When cellulose is treated with caustic and carbon bisulfide the resulting material can be converted into either fibers or sheets. The regenerated cellulose is mixed with a plasticizer and calendered into thin sheets which are

used for wrapping candy, meat, and other foods. Cellulose acetate itself can also be calendered into thin sheets.

Rubber, both natural and synthetic, forms the base material for some derived synthetics. Rubber hydrochloride and chlorinated rubber are two important derivatives. Rubber hydrochloride sheeting, sold as Pliofilm, is used as a wrapping material. It is transparent and waterproof. Rubber, heated in the presence of sulfuric acid, forms a hard, brittle resin which is used as the basis of a cement for bonding rubber to steel and as a base for paint.

Soy beans and other agricultural products have also served as sources of some synthetic plastics. One or two automobiles have been built with plastic bodies for test purposes; it is unlikely, however, that plastics will replace steel in such applications because of their tendency to soften at elevated temperatures.

The Silicones

About 1905 an experimenter in England began an extensive study of certain silicon compounds and some 35 or 40 years later stated that these materials seemed to have no practical usefulness. It happened, however, that several years before this statement was made chemists at the Corning Company had discovered that some of the silicon compounds prepared by the English scientist were highly resistant to heat and water and possessed unusual electrical properties. These were the silanes (silicon compounded with carbon) and the silicones (silicon compounded with carbon and oxygen). From methyl chlor-silane, which is a volatile liquid, there is obtained a vapor which waterproofs any surface exposed to it. This material proved extremely valuable in war-time aviation where it was necessary to maintain perfect radio performance at all times in a plane which might dive from the cold sub-stratosphere into the warm humid atmosphere low over the South Pacific Ocean. Treatment of the electrical parts with methyl chlor-silane prevented equipment failure due to moisture. The reaction of the chlor-silane with water gives resins with varying desirable properties determined by the composition of the basic compounds. For example, because of their high resistance to heat their use as insulation in electric motors makes it possible to build smaller motors and to run them at higher temperatures than

had previously been possible. Other important applications include heat-resistant gaskets to hold the lenses in large military searchlights; and oils for hydraulics fluids, heat exchange fluids and vacuum pumps.

Dyes, Drugs, Insecticides, Soap

No discussion of synthetics would be complete without some mention of other important groups of man-made materials. One of these is dyes. Previous to 1856, dyes were obtained entirely from natural products. In that year an English scientist named William Perkin, while trying to synthesize quinine, discovered how to make a violet dye from aniline which is a coal tar product. This dye, which he named mauve, was the first of the great group of the so-called coal tar dyes that are now available. Today many more and far better dyes can be prepared synthetically than ever came from natural products. Every new synthetic fiber like the ones discussed previously in this chapter requires a complete new set of dyes because of the different chemical structure of the various fibers.

The importance of quinine to the treatment of malaria is well known. When the Japanese occupied the British and Dutch East Indies in 1942 they cut the rest of the world off from about 90 per cent of its normal supply of this drug. Consequently the search for a synthetic substitute or for a method of synthesizing quinine was enormously intensified. Atabrine was the first successful substitute developed. Then in 1945, chemists working at Harvard University succeeded in doing what William Perkin tried to accomplish in 1856, that is, synthesize quinine. Since that time other synthetic drugs have been developed that may prove superior to quinine. Other outstanding examples of drugs known for a long time which are now made synthetically include adrenalin and thyroxin, both formerly obtained only from the glands of certain animals. In addition to synthesizing known drugs, progress has been made toward discovering the particular atomic groupings in the substances that are responsible for certain physiological effects. When this is known it is possible to prepare new substances that have the proper beneficial effects without also producing undesirable ones that the older drug caused. Novocain is an example of this sort of product as also are the sulfa

drugs. It is worthy of note that the active ingredients of penicillin are being discovered and, no doubt, will one day be synthesized.

The production of synthetic drugs for use by man and animals has been paralleled by the production of synthetic chemicals related to plants. The insecticides make up one of the most important groups of such chemicals. Among the most publicized of these is DDT which is too powerful for indiscriminate use on plants because it may also kill pollinating insects. This drug alone is said to have prevented typhus epidemics in Europe during and following the war. It also was an important factor in prevention of malaria among our troops in the Southwest Pacific. A large number of chemicals which are toxic to insects or rodents without being harmful to men and live-stock, are now being produced synthetically. Likewise there are chemicals which kill specific weeds without harming the more useful plants. Still others cause roots to grow on plant cuttings, apples and similar fruits to stay on trees until ripened, and leaves to fall off plants earlier so that certain crops can be harvested more easily.

Another synthetic industry which was born before the war and grew up along with the synthetic rubber industry is the production of synthetic soap. Much of the impetus for its development came from the fact that great quantities of soap are necessary in the production of synthetic rubber by the emulsion process. The products, however, will be available for other purposes and will make it possible to wash clothes, dishes and ourselves readily in hard water or even sea water without the softening and purification processes previously necessary.

We have attempted to summarize in this chapter the highlights as they exist at the present time in a field that is still developing very rapidly, so rapidly in fact that many authorities in industrial organic chemistry felt a year or two ago that the decade might be known in our history as the one which ushered in an Age of Synthetics. Certain events of 1945, of course, now make " Age of Atomic Energy " also a strong contender for this title.

Questions

Discussion

1. Describe the difference between organic and inorganic compounds.

2. Look up the chemical formulas and common names of ten organic and ten inorganic compounds.

3. Name the principal raw materials used in the manufacture of glass.

4. Mention several new uses for glass other than those listed in this chapter.

5. Explain why low-expansion glass should make more satisfactory cooking vessels than would high-expansion glass.

6. Why is carbon so important in organic compounds?

7. Find, from other sources, the structural formulas for two or three organic compounds not mentioned in this chapter.

8. What does one branch represent in the "brush-pile" analogy mentioned in this chapter?

9. Describe briefly the two general types of reaction by which plastics are formed.

10. Look up information on synthetic drugs and dyes other than those mentioned in this chapter.

Multiple Choice

1. The name of Dr. Carrothers is associated with the development of (a) synthetic quinine, (b) heat resistant glass, (c) Celanese, (d) synthetic mauve dye, (e) Nylon.

2. So-called high molecular weight compounds have molecular weights (a) from 0 to 50, (b) from 50 to 100, (c) from 100 to 1,000, (d) from 1,000 to 10,000, (e) over 10,000.

3. List all of the following that are true of any synthetic plastics. (a) It must exist in nature as such. (b) It must be capable of flow under heat or pressure at some stage of its preparation. (c) It must be hard at all temperatures. (d) It must be man-made. (e) It must be of vegetable origin.

4. The phenomenon of edge-lighting is exhibited by (a) all plastics, (b) vinyl resins only, (c) Lucite, (d) Nylon, (e) all thermoplastic compounds.

5. List all of the following that are thermosetting resins. (a) Glass. (b) Vinyl resins. (c) Bakelite. (d) Beetleware. (e) Nylon.

True-False

1. The name of William Perkin is associated with the discovery of aspirin.

2. Temperature has nothing to do with whether a given material is classed as an elastic or a plastic.

3. The molecules of a plastic must be able to polymerize.

4. Modern safety motion picture film is more closely related chemically to the older inflammable film than it is to Pliofilm.

5. The vinyl resins are thermoplastic.

6. Both synthetic elastics and plastics can be classed as high-molecular-weight compounds.

7. Synthetic dyes are used only in cases where a corresponding natural dye does not exist or is not available.

8. Pliofilm is a rubber derivative.

9. Viscose rayon is made from silk.

10. Silicon is one of the ingredients of the silanes.

Chapter 20

FORCE, WORK, AND POWER

For the past several chapters our attention has been focused rather closely on what man has learned about the "stuff" of the world. Modern scientific development, however, cannot be described entirely in terms of materials. What are popularly known as the "wonders of modern science" involve also the quantity that we call energy. In the automobile of today there are incorporated many remarkable achievements in the development and use of substances such as metals, plastics, rubber. In spite of these, the automobile would have no reason for existing if it did not also embody a means of transforming the chemical energy in the gasoline into a form that can be used to turn the rear wheels. An electric light bulb, whose glass and filament represent great advances in research on material substances, must also be able to emit light energy if it is to be worth anything. In other words, in order to make the wheels go round in the world of today, man has had to learn a great deal about energy as well as about matter.

It is with energy, then, that we shall concern ourselves in the next few chapters. Our investigation of it will cover such points as its nature, some of the forms it takes, ways in which it can be changed from one form to another, and methods of transmitting it from one place to another. However, before discussing energy as such there are one or two fundamental concepts which must be developed. Any consideration of energy is bound sooner or later to involve the terms force, work, and power. These words, which we use quite naturally in everyday conversation, have definite specific meanings in science, and it is with these that we shall deal in this chapter. We shall find, as a matter of fact, that the exact scientific definitions of the terms have a great deal in common with the everyday meanings that we attach to them.

Force

Definition of Force

Our purposes in this discussion will be adequately served if we define force as simply " push or pull." The fact that the definition contains only words of one syllable may keep it from sounding very impressive, but it covers the ground fairly satisfactorily. Perhaps too often, scientific definitions seem to imply that a one-syllable word should never be used if a six-syllable one can be substituted.

Kinds of Forces

In considering kinds of forces, the first thing to be noted is that when we classify them we do so in terms of their effects. In speaking or thinking about the force of gravity, one ordinarily has in mind the tendency of objects to fall toward the earth, even though he knows that the term includes forces of attraction among all bodies in the universe. Forces that we call mechanical make us think of push or pull being employed to move something, such as a wheelbarrow or a vacuum cleaner, or to deform something, like a coil spring or a football, or both. The name electrical force, causes us to visualize running motors and generators, or lights, toasters, and irons. In other words, one does not think of the force itself but instead, of some effect that it produces. It is on such a basis that forces are divided into the various kinds.

It is just these effects that concern us most, because what we have learned about them constitutes our experimental knowledge of force. In the next section two groups of such experimental facts will be considered, one connected with certain laws advanced by Sir Isaac Newton, and the other concerning facts that we use in the measurement of forces.

Some Experimental Facts About Forces

Facts Described by Newton's Laws. — It should be noted first that the four laws of Newton that we are about to consider are true scientific laws. This means that they describe the manner in which certain general groups of experimental phenomena have been found to occur. They are not laws which have been passed to tell Nature how she should behave. Thev are statements of how it has been found she

does behave under certain circumstances. Since a complete mathematical discussion of them would be beyond the scope of this book, we shall confine ourselves pretty largely to qualitative considerations. That is, we shall ask what the laws are, what sort of phenomena they describe, and whether or not they seem reasonable in the light of common, everyday experience. First we shall follow up the law of gravitation a little further than we were able to do in a previous chapter, and then we shall consider three laws that, as a group, are known as Newton's laws of motion.

Newton's *Law of Gravitation* expresses the facts concerning a certain kind of force, namely, the force of gravitational attraction. It will be recalled that the statement of this law as given in Chapter 2 was,

> *Every two objects in the universe attract each other with a force which is directly proportional to the product of their masses and inversely proportional to the square of their separation.*

Now we can state the law in a somewhat more efficient form, that is, as an equation:

$$\textit{Force of attraction} = K\,\frac{\textit{Mass of 1st object} \times \textit{Mass of 2nd object}}{\textit{Distance apart, squared}}$$

or, as it would usually be written,

$$F = K\,\frac{M_1 \times M_2}{D^2}$$

where K simply represents a number whose value may be thought of as determined by the kind of units in which the other quantities in the expression are stated. It also will be recalled from the previous discussion of the law, that Newton advanced it as giving a reasonable explanation of the facts observed by Kepler, and stated in the latter's laws of planetary motion. Since that time it has received ample experimental verification. One outstanding example concerned the accuracy of the calculations based upon it which led to the discovery of the planet Neptune.

It should be pointed out that when ordinary sized bodies are involved, the actual force of attraction is not very large. For instance,

the gravitational pull between two 200-pound men who are standing a yard apart is less than two millionths of an ounce. So it is not remarkable that we do not feel ourselves pulled this way and that as a result of the gravitational attraction of objects near us.

The first of the so-called laws of motion is known as the *Law of Inertia.* The facts which it describes are exemplified in everyday life in many ways. Consider the behavior of an individual riding in an automobile when the driver suddenly applies the brakes. The passenger, of course, tends to keep right on going. We say he is thrown forward. Actually, he does keep on moving until he is stopped by the exertion of a force of some kind, usually that of his feet pushing on the floor of the car. Or, think of the passenger's behavior when the car starts very suddenly. In that case we say he is thrown backward. Actually he simply tends to remain at rest and does so until some force acts upon him, usually one exerted by the back of the seat. As a final example, consider a dog that is riding on the right hand running board of an automobile which suddenly makes a left turn. The dog tends to continue his motion in a straight line, and will do so, sliding off the running board unless some force acts to keep him on the car. This force ordinarily will be the one he exerts with his feet. These and all similar phenomena are described in Newton's law of inertia which states,

> *Every body persists in its state of rest or motion in a straight line unless acted upon by some outside force.*

All objects possess this property of tending to remain in their present condition, a characteristic which is called inertia. As a matter of fact, the term need not be confined to the physical world. Mental inertia, which can only be overcome by the application of some sort of outside stimulus, is not entirely unknown.

Newton's second law of motion is known as the *Law of Acceleration.* Before stating the law it will be necessary to make our peace with the quantity " acceleration." In a limited sense it may be thought of as that which an automobile acquires when the driver " steps on " the accelerator. In such a case, then, it is an increase in velocity. Strictly speaking the term includes decrease in velocity as well as increase. Thus, pushing down the accelerator gives the car a positive accelera-

tion while applying the brakes gives it a negative one. Quantitatively, acceleration may be defined as "rate of change of velocity," that is, "change in velocity per unit time." In order to see the kind of units that may be used in measuring acceleration, consider the following example.

Suppose an Automobile Starts from Rest and Speeds Up in Such a Way That It Is Moving,

0 miles per hour	at the end of	0 seconds,
4 miles per hour	at the end of	1 second,
8 miles per hour	at the end of	2 seconds,
12 miles per hour	at the end of	3 seconds,
16 miles per hour	at the end of	4 seconds.

Obviously, the above car's velocity is increasing at the rate of 4 miles per hour for every second that it is in motion. In other words, we could say that its

$$\text{Acceleration} = 4 \text{ miles per hour each second, or}$$
$$= 4 \text{ miles per hour per second.}$$

Or, to take another example, any object falling freely through space can be shown to be traveling about 32 feet per second faster at any instant than it was moving exactly one second earlier. Thus, for such a body,

$$\text{Acceleration} = 32 \text{ feet per second every second.}$$

In general then, acceleration can be measured in units of,

(*A length unit*) per (*A time unit*) every (*A time unit*).

Now let us return to the law of acceleration. It follows naturally after the inertia law which is quite qualitative, stating simply that such and such is true *unless* an outside force acts. This second law goes a step farther and tells what happens *when* that outside force does act. It states that,

> *A force is required to produce an acceleration, the acceleration produced being directly proportional to the force producing it, and inversely proportional to the mass of the object that is accelerated.*

The meaning of direct and inverse proportion has been discussed both in connection with the law of gravitation and in Chapter 13 when the gas laws were being considered. Applied to the above-mentioned automobile, the law means that twice as great a force would be necessary to accelerate the same car 8 miles per hour every second; or that the same force would produce in a car twice as heavy an acceleration one-half as great, that is, of 2 miles per hour every second.

In the interests of accuracy and completeness it should be pointed out that the term " velocity " actually involves direction as well as speed. So a change in direction also is an acceleration and, as in the case of the dog on the running board, can only be produced by the application of a force. We have confined our discussion of acceleration to changes in speed since that is the more familiar use of the term.

The third law of motion goes by the name of the *Law of Interaction.* If a rope is tied to a tree and a man pulls on the free end with a force of 15 pounds, the tree may be said to pull back with an equal force. Or if a 25-pound cake of ice is hanging on a spring balance, the ice may be said to pull down with a force of 25 pounds and the spring balance to pull up with the same amount. To the action of a bullet traveling along the gun barrel, there is a reaction of the gun itself in the opposite direction. These and similar phenomena are described in this law which states,

To every action there is an equal and opposite reaction.

It involves the recognition of the fact that a force cannot exist all by itself. Forces must exist in pairs.

Facts Used as Bases for Measuring Forces. — It is true of force, just as of other quantities we have discussed, that if we are to know very much about it we must be able to measure and compare forces. To do this we find that we must make use of units which are defined in terms of some of the above-mentioned effects.

Consider the pound when it is used as a measure of force. If a person lifts a 5-pound weight, we say quite naturally that he is exerting a force of 5 pounds. We have simply assumed a force unit, the pound, which we would have to define as, " a force equal to the gravitational attraction between the earth and a 1-pound weight." We recognize, again without stopping to think about it, that this 5-pound force does

not have to be used in lifting something. One might push against the wall with a force of 5 pounds, that is, a force 5 times that with which the earth attracts a 1-pound weight. In similar fashion we may define an ounce, gram, kilogram, ton or any other force unit with the same name as a weight unit. It will be observed that weighing devices such as equal arm balances operate by balancing the force of gravity for an unknown object against the force of gravity for one or more bodies of known weight.

On the other hand, if we measure a force with a device such as a spring balance, we make use of the effect a force may have in deforming a coil spring. We are able to employ this type of force-measuring instrument because of an experimental relation known as Hooke's Law. This states that, within the limits of elasticity, the deformation produced is proportional to the force producing it. With such scales we usually use the same kind of force units as those described above.

Although there is no force-measuring instrument which is based on the effect a force may have in accelerating an object, there are force units that are so defined. The one in the English system is called the *poundal*. By definition, it is the amount of force that would speed up a 1-pound mass at the rate of 1 foot per second every second. This is the equivalent of about 1/32 of a pound of force. The analogous metric unit is more commonly used and is called the *dyne*. It is the force necessary to accelerate a 1-gram mass at the rate of 1 centimeter per second every second. It amounts to only 1/980 of a gram of force, or as someone has calculated, about the force that would be required to lift an average-size mosquito.

So it seems to be true that while we talk about forces, we think rather of their effects. Then we make use of these effects in various ways, including the development of instruments for force measurement and as bases for defining force units.

Work

Definition and Relation to Force

Before formally defining the quantity "work," let us mention a few operations to which the term would naturally be applied. Everyone would agree that work is being done in all of the following

cases: when a housewife pushes a vacuum cleaner, when a man pushes a lawnmower, when a horse pulls a wagon, when water turns a water-wheel, when a locomotive pulls a train.

Now it may be noted that two elements are common to these and all similar examples. First, a force is being exerted and, second, that force is producing a displacement. In other words, *work* involves *force*, which has already been discussed, and one additional quantity, *displacement*. By actual definition we may say,

$$Work = Force \times Displacement$$

where the displacement must be measured in the direction in which the force acts.

The Measurement of Work

In the previous discussion of force we said that a force of 1 pound is required to lift a 1-pound weight. Now if that weight is lifted through a vertical distance or displacement of 1 foot, work is done. This amount of work is called a *foot-pound*. By definition, then, we can say,

> *A foot-pound is the amount of work done when a force of one pound produces a displacement of one foot.*

The reader will have no difficulty in defining for himself in similar fashion other units in both the English and metric systems, such as ounce-inches, yard-tons, gram-centimeters, kilogram-meters. It will be noted that all the above work units involve force units of the kind that were defined in terms of gravitational attraction.

Obviously though, work is also done if a force of a poundal or a dyne produces a displacement of a foot or a centimeter. Thus the poundal-foot and the dyne-centimeter are perfectly good units of work. The former is seldom used but the latter is a common unit and is given a special name, the *erg*. This, it will be observed, would be the amount of work done if our hypothetical average mosquito were lifted a distance of one centimeter. Since the erg represents a very small amount of work, a larger unit, the *joule*, is also defined. A joule is the equivalent of 10 million ergs.

The Machine as an Aid in Doing Work

The title, "Machine Age," is one of several that are often applied to the period in which we happen to be living. Since machines are devices that man uses to aid him in getting the work of the world done, it is desirable to examine them to some extent in this connection.

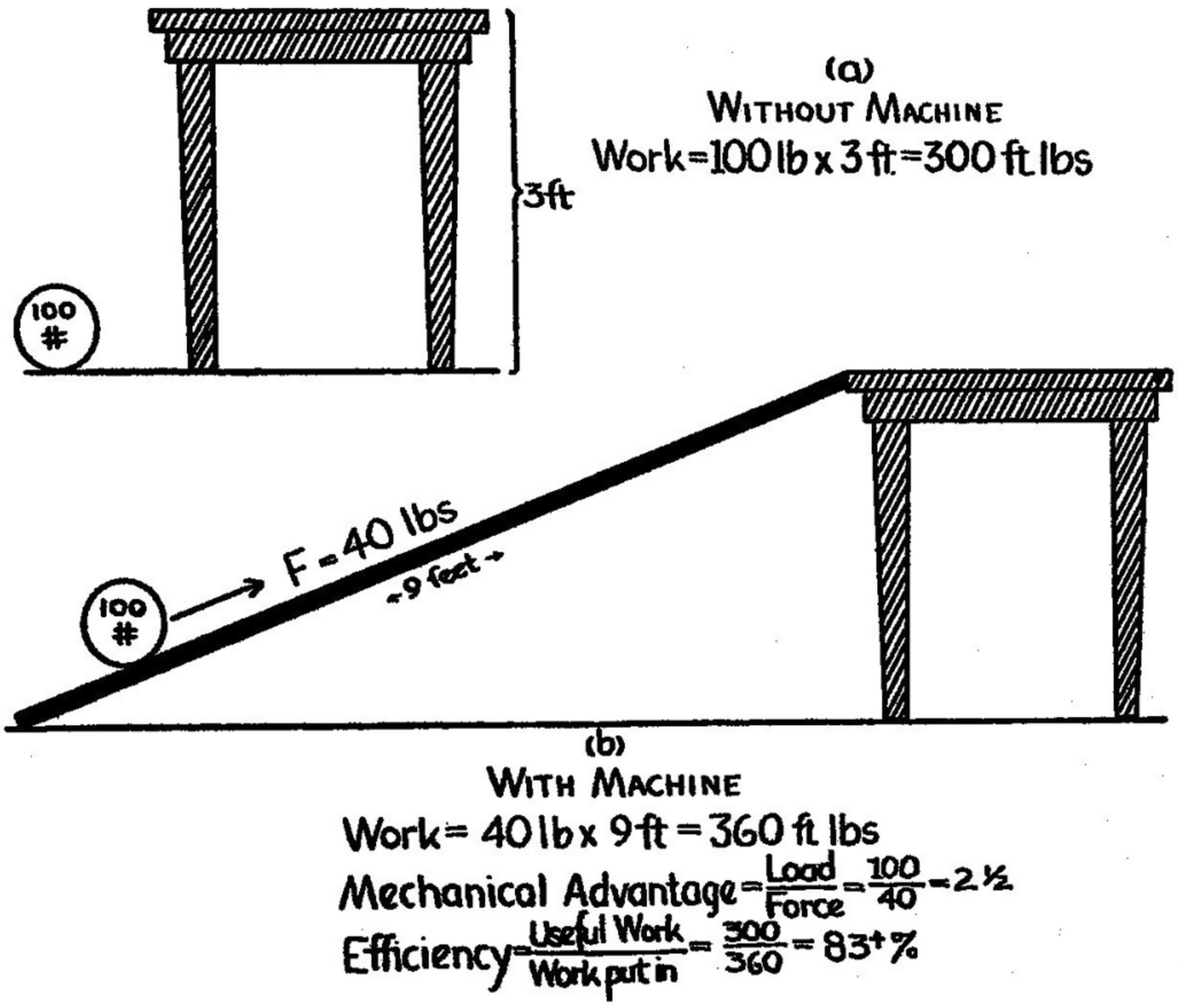


FIG. 106. Diagrams illustrating the use of an inclined plane in raising a 100-lb. weight from the floor onto a table that is 3 feet high.

The first point to be noted is that the problem is greatly simplified by the fact that every machine, no matter how complicated, is a combination of six or less so-called *simple machines*. These six are known as the lever, the inclined plane, the pulley, the screw, the wheel and axle, and the wedge. Since space does not permit our considering all of them in detail, we will examine carefully a specific case of the use of one. With regard to this example we will ask whether it was necessary to use a machine; what, if anything, was gained by its use; and what, if anything, was lost by employing it.

Let us consider the problem of getting a 100-pound keg of nails from the floor to the top of a table which is 3 feet high. First we will think of the work being done without any machine. In that case, the keg is simply picked up and placed on the table. To do this a vertical force of 100 pounds would have to produce a displacement of the keg of 3 feet. Therefore, if no machine is used the work done amounts to 100 pounds times 3 feet which equals 300 foot-pounds, as indicated in (*a*) of Fig. 106. But we might prefer to use an *inclined plane* as an aid in doing this bit of work. In an actual case of this sort a plank 9 feet long was used as the incline up which the keg was rolled. (Part (*b*) of Fig. 106.) In this manner it was found that a force of 40 pounds was all that was required to roll the keg up the plank. This force, however, had to be exerted throughout the entire length of the plank if the keg was to reach the top of the table. Consequently work was done to the extent of 40 pounds times 9 feet which equals 360 foot-pounds, as compared to only 300 foot-pounds required when no machine was employed.

From this example we can establish several fundamental principles that are applicable to all machines. First, it is true, that *the use of a machine always increases the amount of work that has to be done.* The additional work, in part, is expended in overcoming the friction of the machine itself. At first thought then, one may wonder where the advantage lies in using such a device, if more work has to be done with it than without it. One source of gain lies in the fact that it may be possible to exert a smaller force than is otherwise required. In the above case for example, of the two possibilities one would probably prefer to exert the 40-pound force rather than the 100-pound one, even though it did have to be exerted through a greater distance. Then too, very often without a machine forces are so large that the work could not be done. In the above illustration, for example, if the keg had weighed 200 pounds, one man could not have done the work without the machine. But if he used the plank the force required would have been only 80 pounds which he probably could have exerted by himself.

In connection with the use of machines, two important characteristics of these devices must be defined. The first is called *mechanical advantage.* It is a fraction whose numerator is the actual resistance

or load to be moved and whose denominator is the force required to move the load. For the above inclined plane,

$$Mechanical\ Advantage = \frac{Load}{Force} = \frac{100}{40} = 2\tfrac{1}{2}.$$

This simply means that with this particular inclined plane under the conditions existing above, the load which could be moved would be $2\frac{1}{2}$ times as great as the force applied. In the same way, a block and tackle which would enable one to lift a 500-pound safe by applying a force of 25 pounds would have a mechanical advantage of $500/25 = 20$. It should be noted that mechanical advantage has no units. It simply expresses the number of times that the machine multiplies the force which is applied to it. The drawings in Fig. 107 indicate where the loads and forces are applied in various applications of the other simple machines. For each case, the reader might ask himself in just what manner lubrication of these devices would affect, if at all, the quantities of load, force, mechanical advantage, and efficiency. Efficiency will be discussed next.

Efficiency is also a fraction. In this case the numerator is the useful work obtained from the machine, while the denominator is the total work put into it. Making the application to our example, in order to get the 300 foot-pounds of useful work done, that is, to get the keg from the floor to the top of the table, it was necessary to do 360 foot-pounds of work on the machine. Consequently, for that particular device under the conditions used,

$$Efficiency = \frac{Useful\ Work}{Work\ Put\ in} = \frac{300}{360} = .833 = 83 + \%.$$

In other words of all the work put in, a little over 83 per cent was got out in the form of useful work. Although different types of machines vary greatly in efficiency the value always is less than 100 per cent. A device that was 100 per cent efficient would have to be frictionless and therefore would produce perpetual motion. Looking at the page of diagrams of the other machines, it will be obvious that levers should have comparatively high efficiencies, since there are few places where friction can occur. Actually, they may run as high as 98 per cent or 99 per cent efficient under certain conditions. Jackscrews, on the

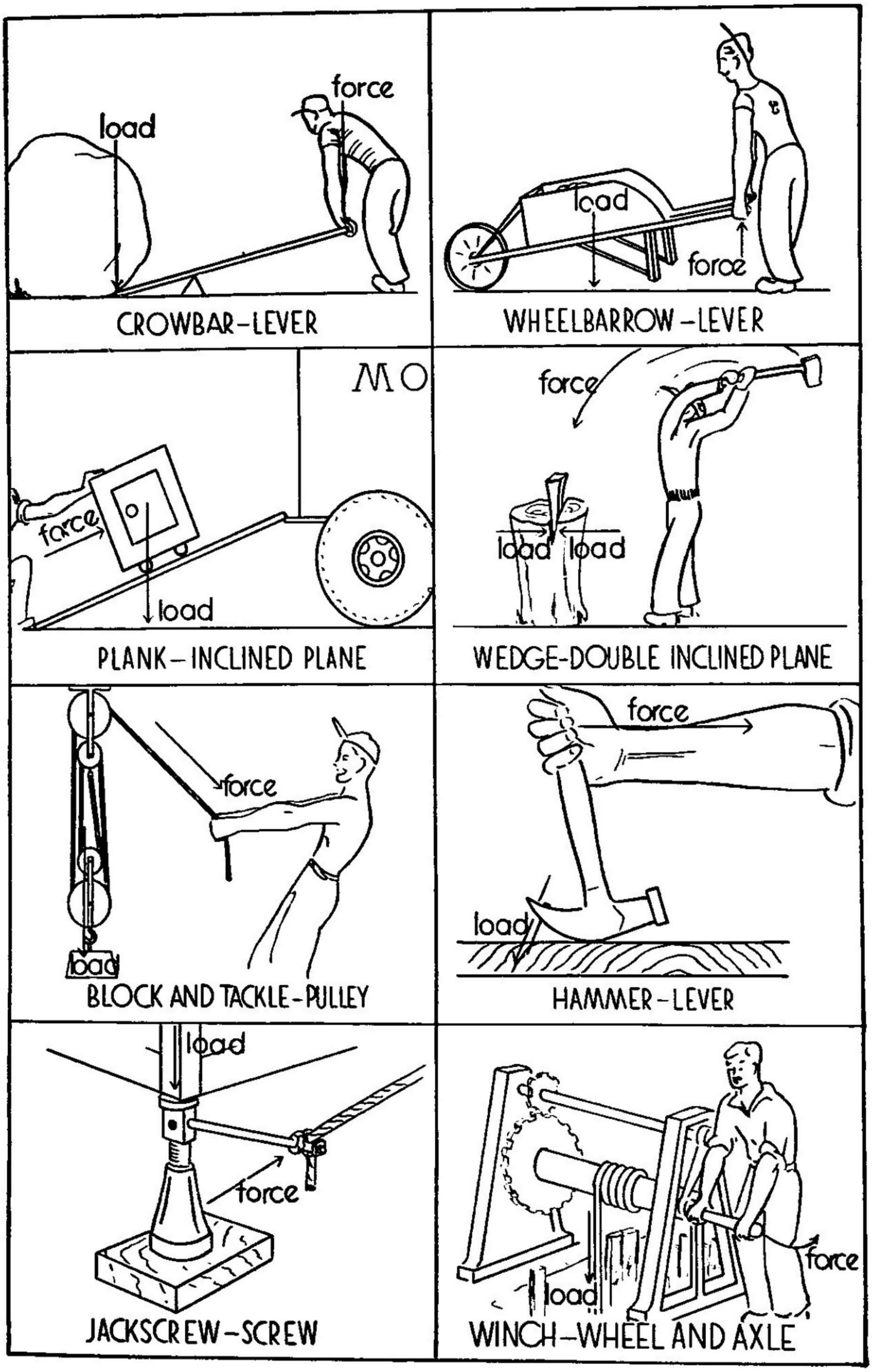


FIG. 107. Diagram illustrating application of several different simple machines

other hand, of the sort used in raising buildings may have efficiencies as low as 15% to 20% because of the enormous sliding friction involved.

Power

Definition and Relation to Force and Work

It will be noted that thus far in this chapter no mention has been made of the time required to do work. In the nail keg experiment no attention was paid to the time spent in rolling the keg up the plank, the work done being entirely independent of that factor. But in many circumstances it is important to know how long a job requires, as well as how much work it involves. The quantity that brings this new factor into the picture is *Power*. It may be defined as,

$$Power = Rate\ of\ Doing\ Work$$

or

$$Power = \frac{Work}{Time}$$

or, since $Work = Force \times Displacement$, we may say,

$$Power = \frac{Force \times Displacement}{Time}$$

So we see that power involves one factor in addition to work, just as work involved one more than did force.

The Measurement of Power

Most of the units in which power is measured follow naturally from the above definition. We have seen that the foot-pound is a unit of work. Obviously then, " foot-pounds per second " would express a rate of doing work, that is, would be a unit in which power could be measured. Similarly, " foot-pounds per minute," " yard-tons per hour," " gram-centimeters per second," " kilogram-meters per minute " would all be acceptable power units. In other words, the quantity power can be measured correctly in terms of,

(*Any work unit*) per (*Any time unit*).

There are in addition to these, several power units with special names. These include the horsepower, the watt, and the kilowatt.

The *horsepower* is simply defined as a rate of working, such that 550 foot-pounds are done every second. This means a 1 h.p. engine

can do 550 foot-pounds of work every second, a ¼ h.p. engine can do work at one-fourth this rate and so on.

The *watt* is a rate of working equal to a joule per second, the joule having already been defined. The *kilowatt* equals 1,000 watts. This last unit is connected with the kilowatt-hour, a term with which the reader will be more or less familiar if he has anything to do with paying electric bills. It should be pointed out that the kilowatt-hour, or KWH, is a unit of work or energy, although it is frequently incorrectly thought of as being a measure of power. It represents the amount of work that is done or the energy consumed in 1 hour by a device which works at the rate of 1 kilowatt. A 1,000-watt electric light bulb burning for 1 hour, or a 500-watt bulb burning for 2 hours, or a 100-watt bulb burning for 10 hours would use 1 KWH of electrical energy.

The horsepower and kilowatt are related as follows:

$$1\ Horsepower = \tfrac{3}{4}\ Kilowatt \text{ (approximately)}.$$

Using the above relationships, the reader may be interested in proving for himself that when electricity costs 5 cents per kilowatt-hour, a nickel will buy enough energy to raise a 1-ton weight through a vertical distance of about a quarter of a mile.

Now, in the light of these definitions, let us calculate the power in two special cases. First we will again consider the experiment involving the 100-pound nail keg and the 3-foot table, assuming that 6 seconds were required to perform the 360 foot-pounds of work. The power then is given by,

$$Power = \frac{Work}{Time} = \frac{360}{6} = 60 \text{ foot-pounds per second,}$$

which would be an entirely correct expression of power, or rate of working. If we wish to express the result in horsepower, since 1 horsepower equals 550 foot-pounds per second, the above rate of doing work is,

$$\frac{60}{550} \text{ or } .11 \text{ horsepower.}$$

In other words the rating in rolling the keg up the plank in 6 seconds would be a little over .1 horsepower.

As a second example, consider the case of a 200-pound man who climbs the stairs to the top of the Empire State building, a vertical height of about 1,250 feet, in half an hour. In getting himself from the street level to this high point he has done work given by,

$$\begin{aligned} Work &= Force \times Displacement \\ &= 200 \text{ lbs.} \times 1{,}250 \text{ ft.} = 250{,}000 \text{ ft.-lbs.} \end{aligned}$$

Since this work was performed in half an hour, which equals 1,800 seconds, his rate of working amounted to,

$$Power = \frac{Work}{Time} = \frac{250{,}000}{1{,}800} = 139 \text{ ft.-lbs. per second,}$$

or to express it as a fraction of a horsepower,

$$= \frac{139}{550} = .25 = \tfrac{1}{4} \text{ horsepower (approximately).}$$

Table 21 gives the quantities discussed in this chapter together with their definitions and units of measurement. The quantity energy, which is last in the table, is considered in the next chapter.

TABLE 21. DEFINITIONS AND EXAMPLES OF UNITS OF FORCE, WORK, POWER, AND ENERGY

Quantity	*Definition*	*Units*
Force	Push or Pull	lb., oz., ton, etc. gram, kg., etc. poundal dyne
Work	Force $\times$ Distance	ft.-lb., oz.-in., etc. gm.-cm., kg.-m., etc. poundal-foot dyne-cm. = erg joule = 10^7 ergs
Power	Rate of Doing Work or $\frac{\text{Work}}{\text{Time}} = \frac{\text{Force} \times \text{Dist.}}{\text{Time}}$	ft.-lb. per sec., etc. gm.-cm. per sec., etc. hp. = 550 ft.-lb./sec. watt = joule/sec. kilowatt = 1,000 watts
Energy	Capacity for Doing Work	Same as work units

Questions

Discussion

1. Define force, work, and power, and state just how each one differs from the other two.

2. How would the force of gravitational attraction between two objects change if both their masses were suddenly tripled? If the distance between them was suddenly tripled?

3. Mention one or two examples of the operation of Newton's law of inertia, other than those described in this chapter.

4. Explain why any statement of acceleration must involve the use of two time units, as "miles per hour per minute" or "feet per second per second."

5. Assuming that it increases at a uniform rate, what is the acceleration of an object that starts from rest and at the end of 3 minutes is going 60 feet per second?

6. If it requires a force of 220 pounds to give a certain automobile an acceleration of 2 miles per hour per second, what force would be necessary to give the same car an acceleration of 1 mile per hour per second?

7. Other things being equal, what force would it take to give a car twice as heavy as the one mentioned above an acceleration of 2 miles per hour per second?

8. Mention some examples of the operation of Newton's law of interaction other than those given in this chapter.

9. Define the following:

a. A pound of force.
b. A poundal of force.
c. A dyne of force.
d. A gram of force.

10. How is the dyne related to the gram? The poundal to the pound?

11. Using some other simple machine, take data analogous to that given in this chapter for an inclined plane and calculate mechanical advantage and efficiency.

12. What is the relation between the dyne and the erg? The erg and the joule?

13. Explain just what advantages there are in the use of machines in view of the fact that they always increase the amount of work that has to be done.

14. What is the mechanical advantage of a machine with which a force of 25 pounds is required to lift a load of 25 pounds?

15. Which one of the following would be the best basis on which to pay a man who mowed your lawn, and why?

a. At so much per force unit.
b. At so much per power unit.
c. At so much per work unit.

16. Decide what data you would need to take in order to calculate your horsepower while running upstairs; then make the experiment.

17. Make the necessary conversions between kinds of units and calculate the value of your effort in running upstairs in Ques. 16 at the prevailing electrical rates in your locality.

Multiple Choice

1. Every body persists in its state of rest or motion in a straight line unless acted upon by an external (a) acceleration, (b) power, (c) force, (d) energy, (e) action.

2. List all of the following that are true when a barrel is rolled up a plank into a truck instead of being lifted up directly. (a) Less work is done. (b) A mechanical advantage of less than one is obtained. (c) A smaller force is exerted. (d) An efficiency of more than 100 per cent is obtained. (e) The power is probably less.

3. Which one of the following might be a correct expression of a machine's mechanical advantage? (a) 6, (b) 6 pounds, (c) 6 foot-pounds, (d) 6 foot-pounds per second, (e) 6 per cent.

4. Which one of Newton's laws would be involved most directly in an explanation of the fact that when an aviator pulls out of a power dive, the blood tends to drain from his head? (a) Law of gravitation. (b) Law of interaction. (c) Law of acceleration. (d) Law of inertia.

5. The force of gravitational attraction between a 10-pound weight and a 2-pound weight 2 feet apart is — (a) 1, (b) 2, (c) 3, (d) 4, (e) 5 — times as great as that between a 5-pound and a 2-pound weight, also 2 feet apart.

True–False

1. In the fraction which represents the mechanical advantage of a machine, both numerator and denominator are forces.

2. A machine which lifts a 55-pound weight through a vertical distance of 10 feet in 1 second is working at a rate of 1 hp.

3. Only objects at rest possess inertia.

4. In addition to the quantities involved in work, a definition of power must also take into account acceleration.

5. In the fraction which represents efficiency, the numerator is never larger than the denominator.

6. To every acceleration there is an equal and opposite acceleration.

7. If it requires 10 seconds for a 1-pound object to fall from a certain balloon to the earth, a 10-pound object would fall the same distance in 1 second.

8. By using a jack to raise an automobile, a man does less work than he would do if he were able to take hold of the car and lift it.

9. A spring balance measures forces in terms of their effects in causing deformation.

Chapter 21

ENERGY AND ITS TRANSFORMATIONS

Definition and Relation to Work

In the preceding chapter we saw that the formal definitions of force, work, and power agree rather well with the meaning we associate with those terms in everyday usage. The same thing is true of "energy." An individual whom we describe as having a great deal of energy is one whom we think of as being able to accomplish a lot. Very much the same idea is involved in the scientific definition which states,

Energy is capacity for doing work.

So a wound watch spring, a boiler under full head of steam, a rapidly moving fist, a person who has eaten regularly, all have a capacity for doing work, or in other words, all possess energy.

Kinds of Energy

Two ways in which energy is classified are of importance to us in this discussion. The first one that we will consider is a grouping based primarily on the kinds of energy effects, or phenomena, that we happen to observe, rather than on any characteristic of the object possessing the energy.

Classification into Mechanical, Heat, Chemical

In speaking of energy in any given case as belonging to one of the above groups we really have in mind the effects of the energy. The situation is much the same as that mentioned previously with regard to various kinds of forces. The important thing about the several kinds of energy is the fact that one can be changed into another *almost* without qualification.

Let us fix our attention on a certain " chunk " of energy that left the sun and arrived in western Pennsylvania several hundred million years ago. While making the journey from its solar home to the earth it was in a form that we would call *radiant energy*. Arriving here, it was absorbed by the trees, mosses, and ferns, of that early period and became *chemical energy*. As such it remained while the vegetation underwent the transformations discussed in Chapter 16, which resulted in turning it into coal. Then let us assume that the coal containing this particular portion of energy was recently mined and shipped to an electric light and power plant to be used as fuel for the operation of a steam turbine. When it is burned the energy emerges in the form of *heat energy*. Used to generate the steam which causes the turbine to operate, we next see it as *mechanical energy*. In this form it is employed in rotating the armature of an electric dynamo and so is changed into *electrical energy*. As electrical energy it is sent out over transmission lines and may be turned into a number of forms. If it happens to be used to ring a doorbell or a fire alarm, part of it becomes *sound energy*. The portion of it that may be put to use in running a vacuum cleaner, lathe or other device employing an electric motor goes back into mechanical energy again. That part that may be used for lighting becomes *light energy*. If some of it finds a career in charging storage batteries, it returns to the form of chemical energy. So it appears that what we call energy can take a number of different forms and can be converted from one to another pretty much as happens to suit our fancy. In the light of this fact, the reader may have begun to wonder whether we can ever create energy or destroy it, or whether all we can do with it is to change it about from one kind to another. This question comes in for more serious consideration a little later in the chapter, in connection with a very important energy law.

Classification into Potential and Kinetic

Here we have energy classified on a different sort of basis, one which concerns certain definite characteristics or properties of the object possessing the energy.

First, we will consider the conditions as a result of which a body is said to have *potential energy*. To begin with the name itself implies something that is hidden or " wrapped up," at least for the time

being. Commencement speakers often refer to the " great potentialities " in the graduating class, meaning possibilities for future action. Much the same idea is embodied in the term when applied to energy.

As a first example let us consider the 5-pound weight shown in Fig. 108. In its position on the floor (*a*) it has no capacity whatever for doing work, that is, it has no energy. When placed on the edge of

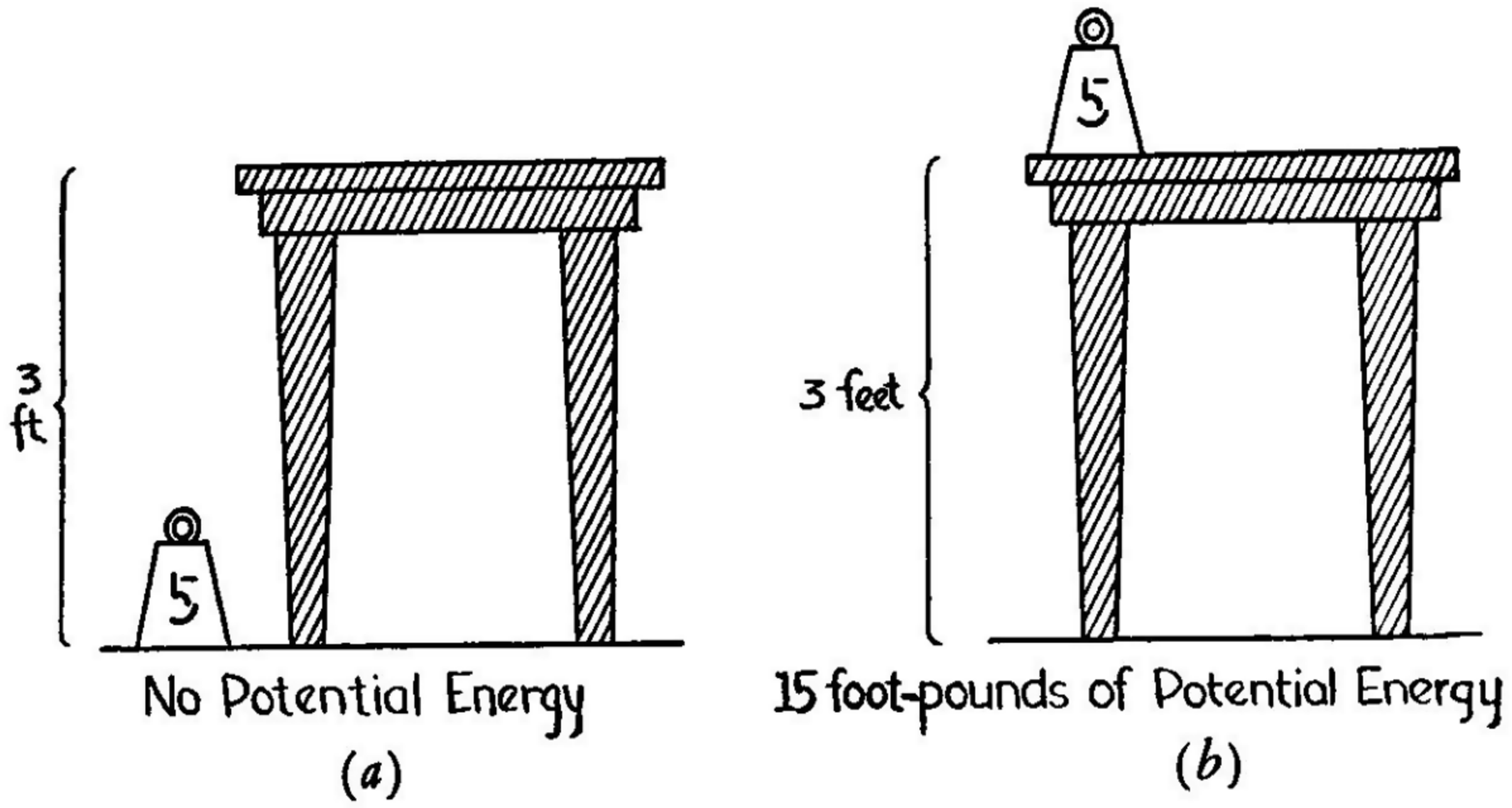


FIG. 108. Diagrams illustrating one form of potential energy, that of position.

the 3-foot table (*b*), it does possess energy. It could be made to do some work by allowing it to fall from the table to the floor. It now possesses energy that it did not possess in the first case, and *it has this energy because of its position.* This is one form of potential energy. In deciding how much this energy amounts to, we must look at the process by which the weight acquired it. Obviously the energy was acquired because someone did work equivalent to lifting the weight from the floor to the table, which according to the relations developed in the last chapter, would amount to " 5 pounds × 3 feet " or 15 foot-pounds of work. Consequently, the weight on the table is said to possess 15 foot-pounds of potential energy that it did not have on the floor. Or, in general, for cases of this sort we can say,

$$Potential\ Energy = Weight \times Height.$$

Energy of position is not the only kind of potential energy. The case of a stretched spring is another example. Here the object is capable of doing work because it has been put in a strained state. As in the above case, the amount of energy it has is equal to the work done on it to put it in its " energy-possessing " condition. When one winds

a watch, one is storing up potential energy in the spring. This potential energy comes back out again as the spring " unwinds " and moves the hands.

A third sort of energy that is commonly spoken of as potential, is the kind possessed by gasoline or dynamite. This is chemical in nature and gives the material the capability of doing work at some future time.

In general we can say that,

Potential energy is energy that a body possesses because of its position, its state of strain, or its ability to do work by chemical reaction.

The above cases still do not cover all the possibilities of obtaining work from an object. Work can also be done by an object purely because it is in motion. This kind of energy is called *Kinetic Energy*. A baseball bat at the instant it hits the ball, an automobile at the instant it strikes another one, a hammer just when it hits the nail, are all examples of cases in which work is done because of motion rather than position, state of strain or chemical reaction. So by definition we may say,

Kinetic energy is energy that a body possesses because of its motion.

The relationship by means of which kinetic energy may be calculated is somewhat more complicated than the one given for potential energy and will not be derived here. Its form is,

$$Kinetic\ Energy = \tfrac{1}{2}\ Mass \times \overline{Velocity}^2$$

which means that the kinetic energy a body possesses increases with the square of its velocity. An automobile traveling 40 miles per hour has *four* times the capacity for doing work of one going 20 miles per hour, or one moving 60 miles per hour has *nine* times that of the 20 miles per hour vehicle. Like potential energy, the units in which kinetic energy is measured are the same as those used for the measurement of work.

In our consideration of the first manner of classifying energy, we saw that conversion could be made from one kind to another. Similarly, it is true that potential energy can change into kinetic and kinetic into potential. This can easily be shown by considering a little

further the previous example of the 5-pound weight on the table. It was pointed out that this object, due to its position, is capable of doing work. Since this work will actually be done only if the weight falls from the table, let us ask what energy changes occur when this happens. On the table the energy is all potential, because the object has no motion. As it falls through the air its height above the floor is becoming constantly less while its speed is becoming constantly greater.

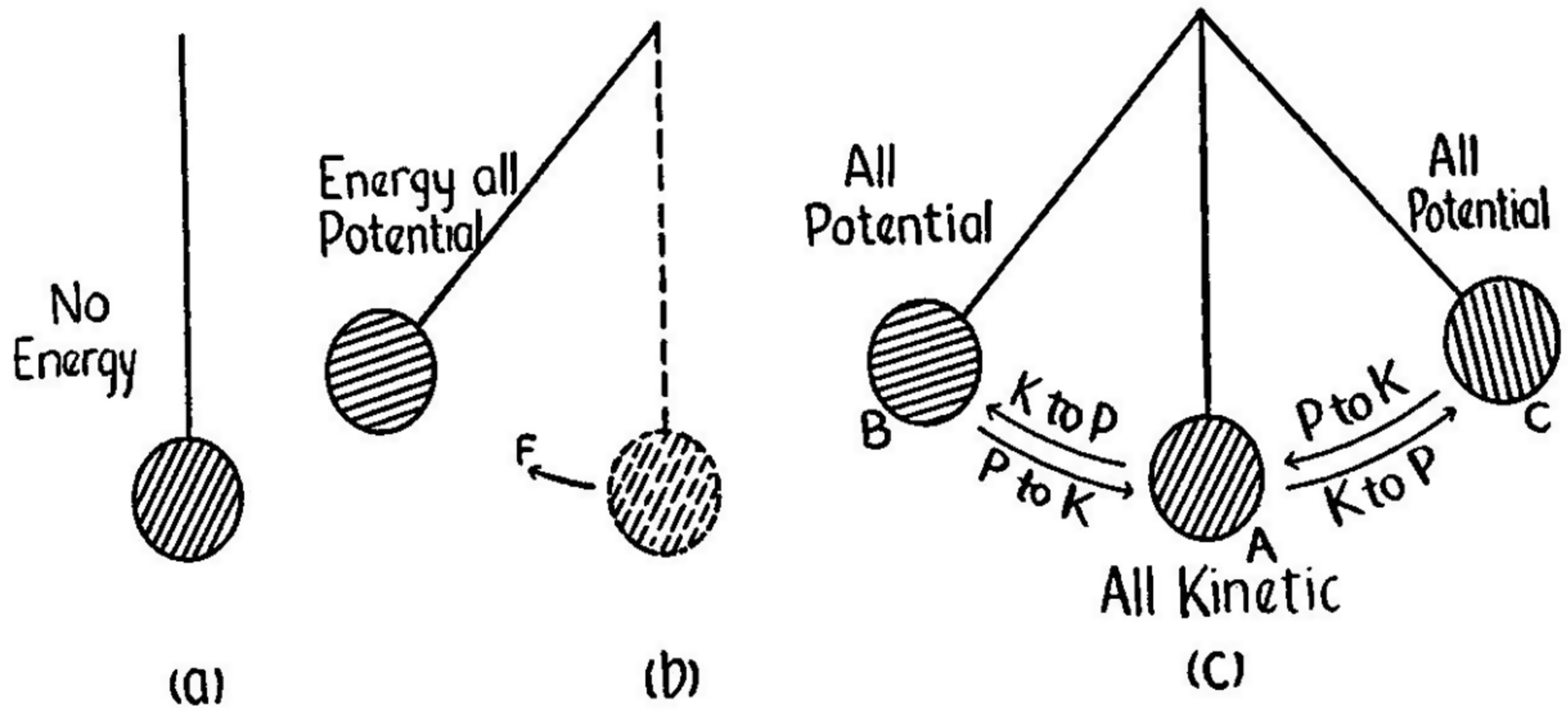


FIG. 109. Diagrams illustrating the acquisition of potential energy by a pendulum and the transformations between potential and kinetic energy as the pendulum swings.

In the light of our definitions of the two kinds of energy, its potential energy is decreasing and its kinetic energy increasing. At the instant it hits the floor, the energy is all kinetic. So potential energy is converted into kinetic energy as the weight falls from the table to the floor. A swinging pendulum such as is shown in Fig. 109 offers an excellent example of repeated energy conversions between these two kinds. Let us assume that the pendulum is found at rest as in (*a*). Someone grasps it and pulls it to the position shown in (*b*). To do this required that work be done. This amount of work is then incorporated in the pendulum when it is in position *B* in the form of potential energy. When released, it swings back and forth as shown in (*c*). During the swing from *B* to *A*, potential energy is changing into kinetic. From *A* to *C*, the conversion is from kinetic to potential. Changes that are the opposite of these occur during the swing back again from *C* to *A* to *B*. The fact that the height to which the pendulum rises is a little less with each oscillation, so that the pendulum eventually comes to rest, makes it clear that during no downward

trip is quite all the potential energy changed to kinetic. Similarly during no rise is there a complete transformation from kinetic energy to potential. This concerns an interesting fact about energy changes which we will take up in more detail a little later in this chapter.

Having considered here two examples of energy conversion between potential and kinetic, the reader will readily see the changes that occur in other cases such as a baseball thrown vertically upward, a screen door that is pushed open and shuts again because of tension in a spring, water going over a waterfall, or a diver moving up and down on a diving board.

The Law of Conservation of Energy

Statement of the Law

For the background of this important scientific principle, we must return to the caloric theory of heat which was discussed in Chapter 14. In that hypothesis heat was regarded as an actual material substance, and an important part of the theory was a law to the effect that heat could be neither created nor destroyed. In other words, the theory involved a "Conservation of Heat" principle. Then there occurred the experiments of Count Rumford and Sir Humphry Davy which could not be explained on the basis of the caloric theory and in which heat definitely was created. Finally, it will be recalled, there was the work of James Joule which resulted in the abandonment of the caloric picture and the adoption of the view that heat is simply a form of energy. It will now be recognized that it is in the form of kinetic energy of the molecules of the material. The work of Joule showed that an exact quantitative relationship exists between mechanical work and heat. This relation, known as the *Mechanical Equivalent of Heat*, states that 778 foot-pounds of work are required to produce 1 B.t.u. of heat. The fact that repetition of this experiment in various forms and performance of many similar experiments always has given the same result led to the expression of the Law of Conservation of Energy. This law states,

> *Energy can be neither created nor destroyed; it can only be converted from one form to another.*

We now have the conservation principle applied to energy as a whole instead of to only one kind, as in the conservation of heat version. A companion principle—the Law of Conservation of Matter—states that matter can neither be created nor destroyed. For almost all practical purposes both of these laws can be thought of as still holding today. It should be pointed out, however, that modern research in atomic physics has demonstrated experimentally that under certain extreme conditions matter can be changed into energy and energy into matter. The idea that there might be a definite, quantitative relationship between matter and energy was advanced by Albert Einstein, in 1905, on purely theoretical grounds. The Einstein equation which relates these quantities, its experimental verification and application in the atomic bomb, and some of the general implications of recent atomic energy studies are discussed in some detail in Chapter 26.

The Sun as the Source of Our Energy

If, as the above law states, we can neither create nor destroy energy, we may well wonder just what its source is. It happens, as was mentioned in Chapter 2, that the ancestry of almost any bit of energy one examines can be traced back to the sun. One such example has already been followed through in the opposite direction in the early part of this chapter. As another illustration, consider the energy the reader uses in turning the pages of this book. He receives it from the food he eats. This food may be traced back to green plants either directly or indirectly. That is, the reader's food is either animal or vegetable in nature. But the animal part itself comes from either vegetables or other animals, and so on until an animal is reached which eats only green plants. Green plants have the ability to absorb energy directly from sunlight. The reader may find it of interest to take other specific examples of energy and trace them back to the star member of the solar system. The only exceptions are ones of minor importance in which the energy came from some star other than the sun. An example of this would be the light from Arcturus, mentioned previously, which was used to activate a photoelectric cell and so set in motion the 1933 Chicago World's Fair. So we can trace the earth's energy back to the sun. Some of the theories that have been advanced to account

for the source of the sun's energy were discussed in the latter part of Chapter 2.

An Amendment to the Conservation Law

We have considered at some length the matter of the conversion of energy from one kind to another. Nature seems to have placed one definite qualification on these conversions, a qualification that may be thought of as a sort of amendment to the Law of Conservation of Energy This amendment becomes quite evident if we consider a few examples. Let us look again at the previously discussed case of the swinging pendulum. We have already referred to the fact that in spite of the energy changing back and forth between potential and kinetic, the pendulum slows down and eventually comes to rest. At first thought this might appear to mean that energy is actually being destroyed, and so constitute a violation of the conservation law. But careful investigation shows that the facts are rather those pictured in Fig. 110. All the potential energy that the pendulum possessed at *B* is not converted to kinetic as it falls to *A*. Some of it is changed to heat energy, due to friction at the point of suspension, and with the air. Then in rising from *A* to *C*, kinetic energy changes back to potential again. But no such reversal occurs as far as the heat energy is concerned. In fact, still more goes into heat. As the pendulum continues to oscillate there is a continual net flow of energy into the form of heat until eventually the pendulum comes to rest. The point to be noted is that at no time does any of this heat energy return naturally to either the potential or kinetic forms. Or we may consider the case of a baseball that is thrown vertically upwards. As it leaves the hand its energy is kinetic. On the way up this kinetic changes to potential and to heat, the latter due to friction with the at-

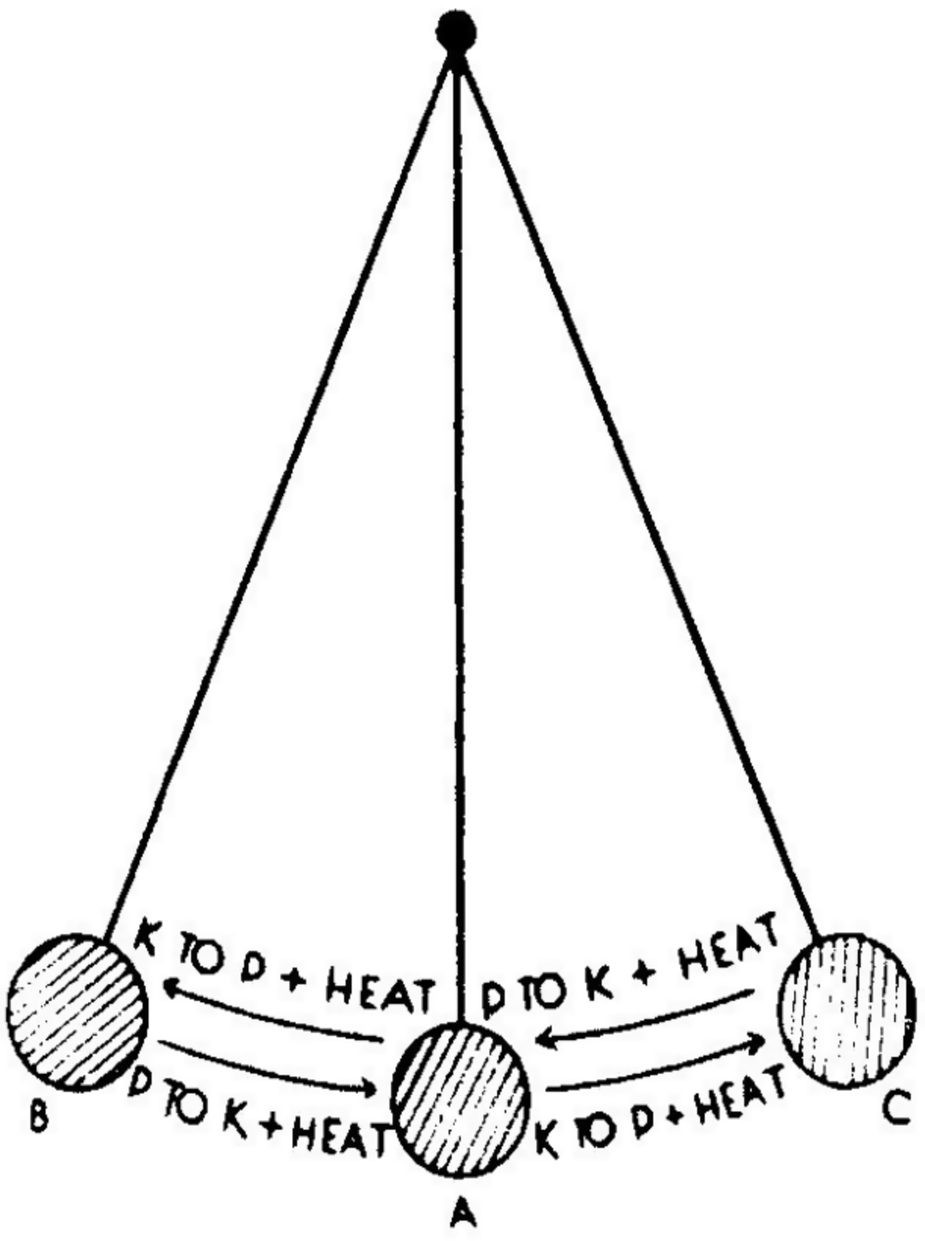


FIG. 110. The complete energy transformations that occur in a swinging pendulum.

mosphere. On the way back down there is a reversal in direction of the kinetic-potential conversion, but the change into heat energy still goes on. In like fashion one might examine all the energy changes mentioned in the first part of the chapter when radiant energy from the sun was followed through a series of reincarnations until it became sound and light. With every change, some energy was transformed into heat. In other words, in Nature's energy family, Heat seems to be the favorite child, always collecting for itself a percentage of the energy involved in any of the transactions among Mechanical, Electrical, Magnetic, Sound, Light, or any of the other children.

This amendment to the Law of Conservation of Energy may be stated,

> *Energy never changes naturally from heat to any of the other forms.*

These facts are of considerable importance to us in this discussion, because, as we shall see in the next chapter, much of our modern use of energy involves making just this conversion from heat to another form, and since it does not occur naturally, we must devise machines which will make it occur.

In order to link our discussion with more conventional treatments of the subject, it should be pointed out that the experimental facts which we have been calling an amendment to the Law of Conservation of Energy, really represent part of the data upon which a law, known in physics by the somewhat formidable title of the Second Law of Thermodynamics, is based. The Conservation Law itself is the First Law of Thermodynamics.

Questions

Discussion

1. Trace the energy you use in turning the pages of this book, back to the sun.

2. Mention circumstances in which the following energy changes might occur:

a. Radiant to chemical.
b. Mechanical to electrical.
c. Electrical to heat.
d. Mechanical to heat.
e. Potential to kinetic.
f. Kinetic to potential.

3. Mention examples, other than the ones given, of objects which possess potential energy and of ones which possess kinetic energy.

4. On the basis of the kinetic theory, what relationship would you say exists between heat energy and kinetic energy?

5. If a book possesses more potential energy when on a table than when on the floor, from where does the additional energy come?

6. Explain the statement, " The fact that a swinging pendulum eventually comes to rest is not a violation of the Law of Conservation of Energy."

7. In the caloric theory of heat there was a basic assumption to the effect that heat could neither be created nor destroyed. Mention some experimental evidence that contradicts this statement.

8. Look up some information on the so-called " Heat Death of the Universe " and link it to the discussion in this chapter of what we have called an amendment to the Conservation Law.

Multiple Choice

1. List all of the following that may be said to possess more potential than kinetic energy. (a) A freight car at rest on a level track. (b) The same car in motion on the same level track. (c) A roller coaster car at the top of an incline. (d) A bullet as it leaves the muzzle of the gun. (e) A baseball at the highest point it reaches after having been thrown vertically upwards.

2. Energy may be defined as capacity for (a) exerting force, (b) action, (c) great power, (d) doing work, (e) producing motion.

3. Energy may be (a) created but not destroyed, (b) destroyed but not created, (c) created and destroyed but not transformed into various forms, (d) changed from one form to another but neither created nor destroyed.

4. During which of the following (one or more) is kinetic energy being transformed into potential energy? (a) A roller coaster car going down an incline. (b) The same car going up the next incline. (c) A storage battery while it is being charged. (d) A parachutist falling through the air, before he pulls the rip cord. (e) The same parachutist after he pulls the rip cord. (f) A man winding a watch.

5. If the transfer could be made without loss, the energy necessary to heat a pound of water from 70° F. to 71° F. would be sufficient to lift a 200-pound man through a vertical height of between — (a) 0 and 2, (b) 2 and 4, (c) 4 and 6, (d) 6 and 8, (e) 8 and 10 — feet.

True–False

1. The fact that a swinging pendulum slows down and eventually stops proves that there are exceptions to the Law of Conservation of Energy.

2. If a block slides down an incline due to its own weight only, its actual kinetic energy at the bottom will be less than was its potential at the top.

3. At 80 miles an hour, an automobile possesses 16 times as much kinetic energy as it does at 20 miles per hour.

4. It is correct to say that heat can be neither created nor destroyed.

5. More energy changes naturally from mechanical to heat than from heat to mechanical.

6. A man's physical energy cannot be traced back to the sun without a green plant's being involved somewhere along the way.

7. In a freely swinging pendulum, the total amount of potential energy that changes to kinetic is greater than the total which is transformed in the opposite direction.

8. The statement of the Law of Conservation of Energy is contradicted by the fact that no machine can be 100 per cent efficient.

9. In the light of the previous discussion of the kinetic or molecular theory, heat may be thought of as a form of kinetic energy.

10. Potential energy could be measured correctly in joules.

Chapter 22

HEAT ENGINES

General Description of Heat Engines

In the first part of this chapter we shall consider heat engines in general, defining them and describing their general functions. Then we shall discuss three specific types, the steam engine, the internal combustion engine, and the jet-propulsion engine.

Relation to Previous Discussion

First, let us remind the reader of the statement which, in the preceding chapter, we called an amendment to the Law of Conservation of Energy. That amendment was, "Energy never changes naturally from heat into other forms." Keeping this fact in mind, let us ask just what are the immediate sources of most of the energy used in the world today. Consider, for instance, the energy expended in transportation of all kinds. In its final form it is mechanical energy, but almost all of it can be traced back to the heat energy of gasoline, oil, coal, or other fuel. Similarly, much of our electrical energy comes originally from the heat energy of some type of fuel, fuel which was burned to produce the energy that drove the dynamo. In other words, most of the energy that makes the wheels go around in our modern world is the result of transformations from heat energy to other kinds. Consequently, when we make energy changes of this sort occur we find that there are always large losses involved. Or, to state the same fact quantitatively, although 778 foot-pounds of work are required to generate 1 B.t.u. of heat, we cannot take 1 B.t.u. of heat and make it do 778 foot-pounds of work. In fact, we shall find that under the very best conditions we cannot get much over a fourth of 778 foot-pounds from the 1 B.t.u. of heat. It is this loss, inherent in any energy

transformation from heat to other forms, that is the principal cause of the low efficiency of heat engines.

Definition of a Heat Engine

As might be suspected from the above discussion, a heat engine may be defined as any device which changes heat energy into mechanical energy. Included in this classification are all such machines as steam engines, gasoline engines, Diesel engines, and jet-propulsion engines. Even the human body itself is a heat engine since it is able to do mechanical work as a result of the energy provided by the food that is eaten.

It should be emphasized that, even though heat engines are relatively inefficient, they do make it possible to use fuels to do work which such substances could not be made to do in any other way. Five gallons of gasoline all by itself is not capable of providing much in the way of what we may call controlled transportation. But, if it is run through the kind of heat engine we use in an automobile, 80 to 100 miles of such transportation are made available for the car and its occupants.

The Steam Engine

It has often been pointed out that man is the only animal that makes use of energy other than that resulting from his own muscular efforts. It is also true that even man did very little about external sources of energy until the latter part of the seventeenth century. During all the preceding thousands of years his only such supplementary energy came from domesticated animals, water, and wind. But with the industrial revolution came the need for doing various kinds of work that could not be performed satisfactorily by beasts of burden, water-wheels, or windmills, and so the steam engine was developed.

Steam Power before 1700

Previous to 1600 a number of attempts had been made to harness steam power but without any appreciable success. By 1600, however, changing industrial conditions in England had enormously increased the demand for coal. Since one of the principal hindrances to in-

creased coal production was the water which collected in the mines, there was an urgent need for a power-driven water pump of some kind. The first such pump to have any extensive commercial use was patented in 1698 by Thomas Savery. Its principle of operation was very simple and its applications extremely limited. Steam was allowed to enter a cylinder from which a pipe led to the water in the mine. As the steam condensed, creating a partial vacuum, atmospheric pressure forced water from the mine up into the cylinder. A valve in the inlet pipe was then closed and new steam introduced, forcing the water out through a pipe leading up to the surface and at the same time again filling the cylinder with steam. The cycle was then repeated. Since the engine had no moving parts it could not be adapted for use in running other machines but was strictly limited to pumping water. The Savery engine had many shortcomings and within a few years was supplanted by a device developed by Thomas Newcomen.

Newcomen's Engine

The earliest direct ancestor of the modern steam engine was the " atmospheric engine " which was built around 1705 by Thomas Newcomen (1663–1729). Newcomen's engine, like Savery's, was built primarily for pumping water out of mines. Although the two men were contemporaries and fellow-countrymen, there seems to be no evidence that either one knew about the work of the other.

The essential features of the Newcomen engine are shown in Fig. 111. It marked the first important use ever made of a piston operating in a cylinder. The cycle of operation was as follows: Valve *A* was opened and the cylinder filled with steam. This steam condensed, creating a partial vacuum in the cylinder. In this device cold water was sprayed into the cylinder through the jet *B* to speed up the condensation process. Because of the lowered pressure within the cylinder, atmospheric pressure forced the piston down. The downward motion of the piston was communicated through the walking beam and raised the pump rod. Next the valve *C* was opened and the water drained out of the cylinder. Opening this valve also permitted air to enter the cylinder and the weight on the left end of the walking beam dropped, raising the piston up to its original position and lower-

ing the pump rod again. Although the valves were opened and closed by hand in the earliest models, by 1712 Newcomen engines were in use in which this work was done automatically by levers attached to the moving parts of the machine.

One of the principal advantages enjoyed by the Newcomen engine over the Savery was the fact that steam pressure in the later engine was not much greater than that of the atmosphere. For this reason the danger of explosion was considerably reduced. Also the engine produced an up and down movement of a beam, and so possessed possibilities for doing other things than pumping water. Actually not much else was done with it until after certain improvements had been added by James Watt.

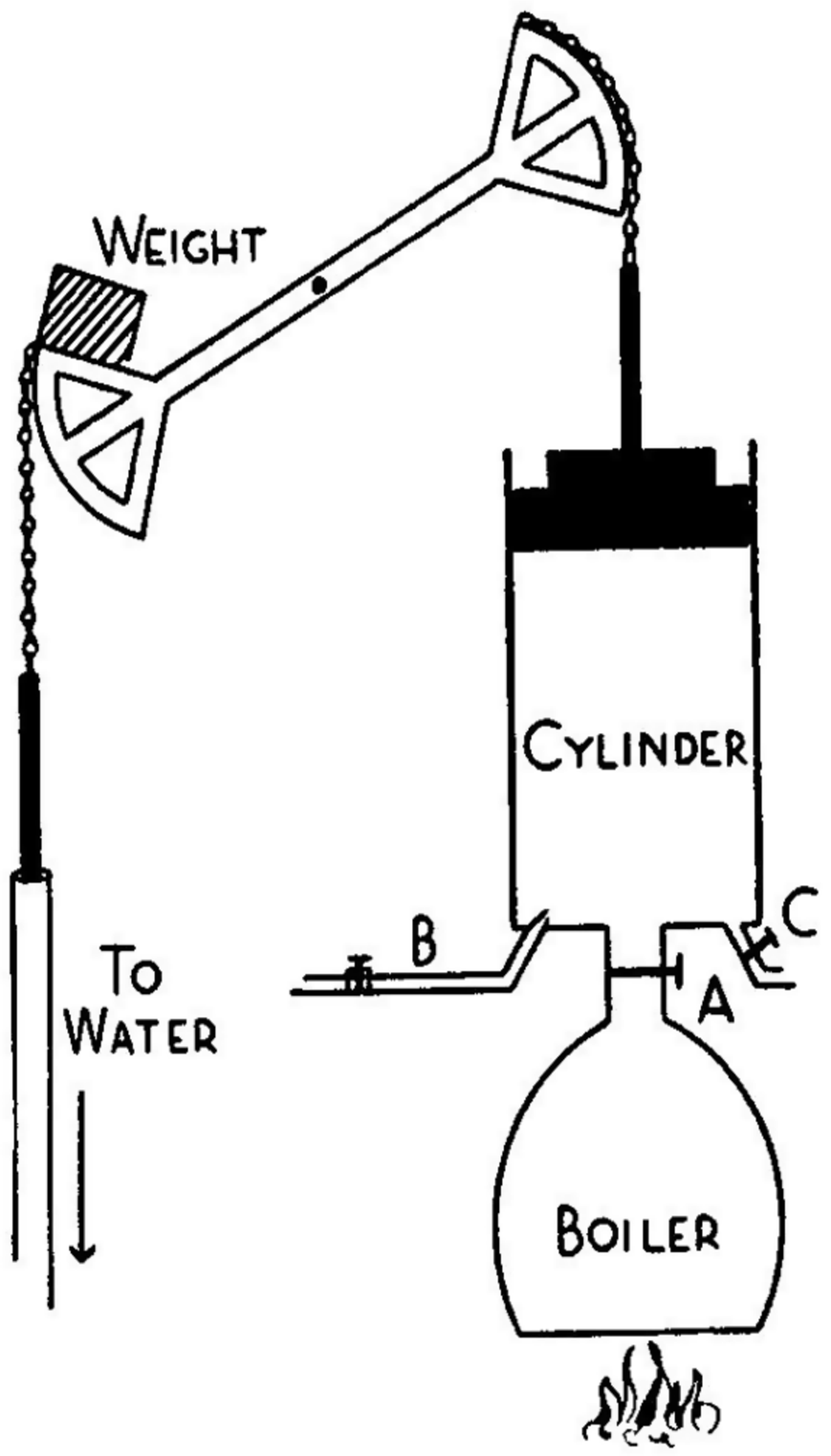


FIG. 111. Schematic diagram of the steam engine of Thomas Newcomen — 1705.

Watt's Engine

James Watt (1736–1819) is frequently credited with having been the steam engine's inventor. Almost everyone has heard how Watt as a boy was so impressed with the power exhibited by steam as it raised the lid of his mother's teakettle that he grew up and invented the steam engine. As a matter of fact this tale belongs with some of the other fables we have mentioned because, as we have seen, steam engines were being used industrially some years before James Watt was born. What Watt did do was to add a number of improvements to the Newcomen engine, improvements that were so important and revolutionary that it is not surprising the whole invention has so often been ascribed to him. We will consider four of his improvements that were of outstanding importance.

First, he noticed that the Newcomen engine seemed to require an enormous quantity of steam in order to do a rather small amount

of work. A little experimentation convinced Watt that this was because the entire piston and cylinder cooled down with each condensation of steam and so had to be heated up again after each power stroke. In fact he found that about seven-eighths of all the steam generated was being used in this fashion and only one-eighth going into the production of useful work. To remedy this he added a small separate chamber that was kept at a low temperature and in which the steam was allowed to condense. Since the condensation of the steam created a partial vacuum in both chambers, the engine could operate as before. Now the cylinder could be kept warm all the time, eliminating the necessity for constantly re-heating it.

Second, Watt increased the pressure at which the steam was admitted to the cylinder. By doing this he was able to use the steam pressure to raise the piston, instead of depending on the counterweight on the walking beam. At this point in the development the down stroke was still accomplished by the creation of a partial vacuum below the piston.

The third improvement consisted in making the engine double-acting. This means that he devised a system of tubes and valves so arranged that steam could be admitted alternately on either side of the piston. Then steam pressure was used both to force the piston up and to drive it back down again, eliminating the use of atmospheric pressure.

The fourth important feature that Watt added resulted from his observations of the exhaust steam that issued from the engines. He noticed that it was at a fairly high pressure and reasoned from this that it still contained a fairly large amount of energy that was being wasted. Up to that time steam had been admitted to the cylinder during the entire stroke of the piston. Watt changed this so that the steam was cut off after the piston had traveled only part of the way along the cylinder. The force for the rest of the trip was exerted by the expansion of the small amount of steam that had gained entrance. In Watt's engine the steam cut-off occurred at such a time that the steam expanded to three or four times its original volume during the rest of the stroke. In some modern engines this expansion is as much as two hundredfold.

One more improvement must be mentioned here, although it cannot be credited to Watt. It concerns the type of motion obtained from the engine. It will be noted that both the Newcomen engine and the Watt engine, as discussed thus far, produced only "up and down" motion. Such motion is called "reciprocating," a term that includes all motions that we might describe as up and down, back and forth, or to and fro. However, it is obvious that most of the things for which we use engines today involve making wheels, gears, and shafts go around rather than in causing something to go back and forth. In other words, an engine which delivers rotational motion is a much more valuable device than one which generates only reciprocating motion. But any engine which has a piston operating in a cylinder produces reciprocating motion of that piston, so that if we are to get rotational motion out of it, the engine must contain some device which changes the one type of motion into the other.

The device which we use today to change the reciprocating motion of a piston into rotational motion of a shaft was first applied to steam engines around 1780. It is called a crankshaft and its operation may be illustrated by the familiar brace and bit used by carpenters to bore holes in wood. It will be recalled that when this device is being used, the carpenter's hand and arm move back and forth but the bit rotates as it goes through the wood. The brace represents the crankshaft and the carpenter's hand and arm the connecting rod. The offset part of the brace which is grasped by the hand is called a "throw." If the reader will look back at the Newcomen engine pictured in Fig. 111 he will see that one might replace the pump rod and chain on the left side of the walking beam with a connecting rod and fasten it to the throw of a crankshaft and so make the shaft rotate as the engine operated.

By the time the above five improvements, together with many others of less importance, were added to Newcomen's engine its efficiency had been increased six or eight times. Although all of Watt's engines looked like Newcomen's in that they used vertical cylinders and overhead walking beams, they were quite close in principle and cycle of operation to the slide-valve steam engine of today which we will consider next.

The Slide-Valve Steam Engine

Cycle of Operation. — The essential parts of a simple slide-valve steam engine are shown in Fig. 112. We will trace the steps in its cycle of operation. Steam enters the steam chest from a boiler not shown in the diagram. As the engine is pictured here, the steam moves down through the passage *L* and exerts its force against the left side of the piston, pushing it toward the right. As mentioned above, the steam is cut off before the piston has traveled far so that most of the stroke is caused by the expansion of the first steam to enter the

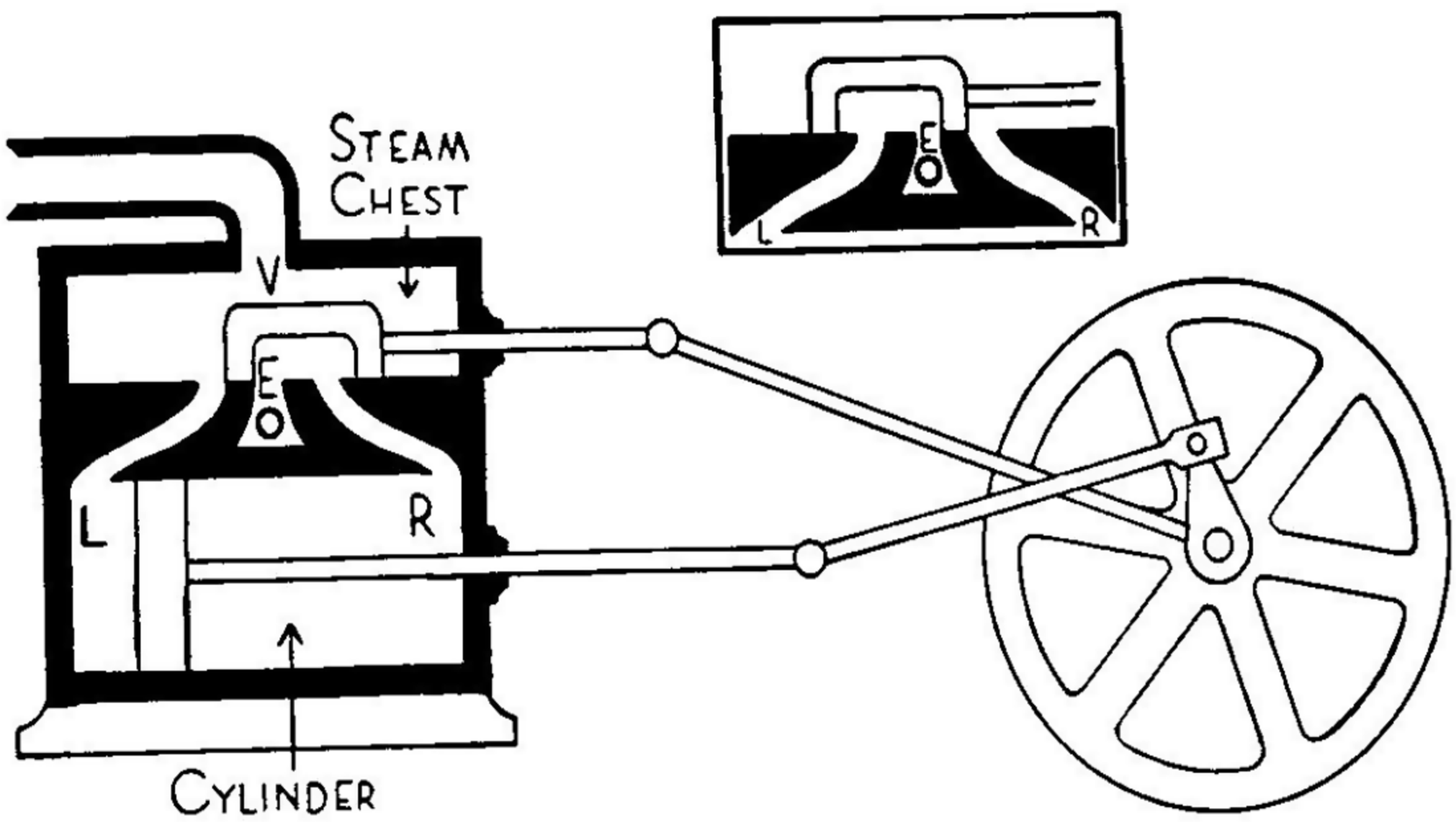


FIG. 112. Cross-sectional diagram of a simple slide-valve engine when the piston is at the left end of the cylinder starting back.

cylinder. This cut-off is accomplished by the valve *V* sliding a little to the left. As the piston moves to the right, exhaust steam from the preceding stroke is pushed ahead of it, up through the passage *R*, and out the exhaust pipe *E*. By the time the piston reaches the right-hand end of the cylinder, the valve has moved to the left to the position shown in the insert. Now steam from the steam chest enters the cylinder through *R* and forces the piston back. Exhaust is now taken care of through *L* and out through *E*. After the piston reaches the left-hand end of the cylinder, completing one cycle, the procedure is repeated. Although modern engines include certain additional features designed to increase efficiency, the above description covers the essentials of their operation.

It should be pointed out that the slide-valve described above is not the only kind of device used to provide for admission and exhaust

of steam in reciprocating engines. However, it is the most widely used, other valve arrangements simply accomplishing exactly the same results by slightly different mechanisms.

Additional Features. — One method of increasing efficiency is to have the exhaust steam go into a condensing chamber. When it is simply exhausted into the open air, atmospheric pressure of about 15 pounds per square inch is constantly pushing back on the piston. If, however, it condenses in a closed chamber, thus creating a partial vacuum, the back pressure on the piston is considerably reduced. Also, the condensed steam can then be conducted back into the boiler and used over again. This effects an additional saving in fuel since this water is already at a comparatively high temperature and so requires less heating.

Higher efficiencies are also obtained if more than one cylinder is used. Steam that has worked against one piston, as described above, still has some energy in it. This can be used by conducting the steam to another cylinder and letting it work against another piston instead of simply exhausting it. In modern engines the steam may be employed in as many as four cylinders before being finally conducted to the condensing chamber.

Efficiency. — The efficiency of the average steam engine is about 18.5 per cent and of the average steam boiler approximately 70 per cent. Thus the total efficiency of the installation from fuel through to useful work averages 18.5 per cent of 70 per cent, or about 13 per cent. By the use of very high initial temperatures efficiencies as high as 24 per cent have been reached in the best modern reciprocating engines but with valve mechanisms other than the slide-valve type.

The Steam Turbine

In the steam turbine, rotational mechanical energy is obtained from the heat energy in a fuel by a procedure somewhat more direct than that just described for the reciprocating type of engine. Two men, an Englishman and a Swede, were chiefly responsible for early turbine development. The former, Charles Parsons, built his first machine for use in running an electric generator. The latter, Carl de Laval, was interested in building a device that could be used as a

source of energy for his newly developed cream separator. Both did their work during the latter part of the nineteenth century.

As might be expected, early steam turbines looked very much like water-wheels and operated by directing jets of steam against flat vanes on paddle-wheels. Later developments which led to increased efficiencies included the use of curved paddles and of directing the

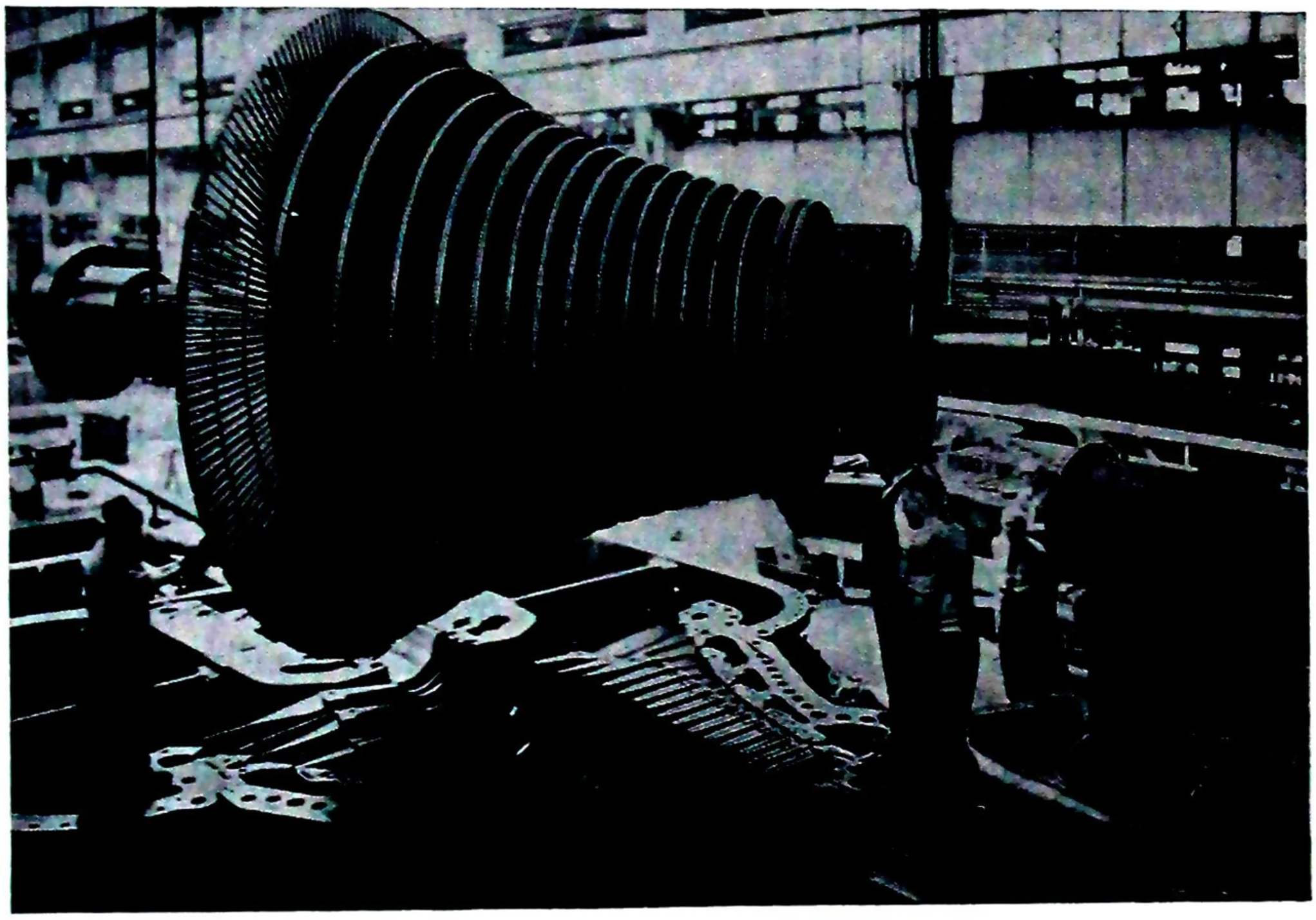


FIG. 113. The rotating member or rotor of a General Electric 75,000-kilowatt steam turbine. (General Electric Co.)

steam against a whole series of wheels instead of only one. The rotating member of a large modern turbine is shown in Fig. 113. The tapered design enables the device to make use of the expansive power of the steam which enters at the small end and is exhausted at the large end. Stationary vanes, which alternate with the rotating vanes in the completed turbine, serve to reverse the direction of the steam between impacts against the rotor.

Steam turbines attain maximum efficiency, between 20 per cent and 25 per cent, only when run at high speed. Although they average a little higher than do reciprocating engines, the maximum efficiencies

of the two kinds are about the same. Turbines, however, can be built in larger sizes than can the other type, some generating more than 160,000 horsepower. They find their principal applications in circumstances where high-speed, continuous, rotational motion is desired. Two important examples of their use are as sources of energy for running electric generators and for operating ocean liners.

The Internal Combustion Engine

Two specific kinds of internal combustion engines will be considered. But first we shall discuss a group of general features which are found in all such engines, introducing them by comparing the internal combustion type of device with the reciprocating steam engine described earlier in the chapter.

Comparison with the Steam Engine

The most important similarity between the internal combustion engine and the steam engine is that both are heat engines. Also, both of them accomplish the conversion of heat energy into mechanical energy by first generating a reciprocating motion of a piston in a cylinder and then converting this into rotational motion by the use of a crankshaft and one or more connecting rods.

The fundamental difference between the steam engine and the internal combustion engine concerns the place in which the fuel is burned. The name "internal combustion" suggests that the burning takes place inside of something. That something is the cylinder. So the steam engine might be called an "external combustion" machine because the fuel is consumed entirely outside the cylinder or cylinders. The rate at which combustion occurs also differs widely in the two types of engines. In the steam outfit the fuel burns comparatively slowly, the heat being used to generate steam. The steam then acts as an intermediate working substance, a sort of "middleman," that conveys the energy into the cylinder and expends it in moving the piston. In the internal combustion engine no such intermediate material is necessary. The fuel burns with great rapidity within the cylinder and the expanding gaseous products of combustion themselves drive the piston.

Essential Features

The very fact that combustion occurs in the somewhat inaccessible location described above introduces certain rather obvious problems that must be solved if such an engine is to operate. We will consider four of the most important of these.

Vaporization of the Fuel. — It is obvious that, with combustion occurring within the cylinder, a solid fuel like wood or coal could hardly be satisfactory, at least not unless it were to be ground into a fine powder. It is conceivable that a liquid might be made to flow into the cylinder but it would undoubtedly burn too slowly to be satisfactory. A vaporized fuel, on the other hand, not only can be made to burn very rapidly, but it also can easily be conducted into the combustion chamber. Therefore, internal combustion engines always use fuels that are in vapor form at the time combustion occurs. If the fuel happens to be purchased in liquid form, then some means of converting it into a vapor must be provided. For instance, in the automobile engine the carburetor has this function. Space here does not permit a detailed explanation of the construction and operation of the carburetor, but one may think of it as working in somewhat the same manner as an insect spray gun or atomizer. That is to say, air moves rapidly past the end of a tube from which the gasoline emerges and mixes with it, the proportion being about 15 pounds of air to 1 pound of gasoline. In this fashion, every gallon of gasoline is mixed with approximately the amount of air that would be in a room 10 feet square and 12 feet high.

Conduction of the Fuel to the Cylinder. — The next step is to get the mixture of fuel and air from the carburetor to the cylinders. This transfer is made through a pipe called the "intake manifold." The vapor is forced through the manifold as a result of a partial vacuum which is created in the cylinder. The manner in which this vacuum is created will be considered a little later. Since combustion occurs more readily if the vaporized fuel has been heated, the manifold runs close to the engine proper where it will be kept warm.

Ignition of the Fuel. — Igniting the fuel presents no particular problem in the case of a steam engine because the coal is readily accessible in the firepot under the boiler. But there is a problem connected

with this part of the operation in an internal combustion engine where the fuel is deep inside the engine itself. It is obvious that the method of igniting the fuel must be an automatic one. A completely satisfactory procedure of this sort had to await the development of the electric ignition system, thus furnishing another example of the interdependence of different fields of technology. In the automobile engine of today the fuel is ignited by an electric spark which jumps across a gap between two wires on the bottom of the spark plug.

Exhaust of Waste Products. — Here again the average steam engine presents no particular problem. If the fuel is coal or wood, the waste products of combustion, that is, the ashes, may be exhausted by means of a man and a wheelbarrow or their equivalent. But in the internal combustion engine, the waste products are inside the cylinders. The first point that suggests itself is that, under such conditions, the easiest kind of waste products to get rid of would be gaseous ones, because they could simply be blown out through an exhaust manifold and pipe. The ideal situation in this type of engine is to have the fuel pure enough and combustion complete enough so that all the products of the burning will be gases. Since this ideal is never quite attained, there are always larger or smaller amounts of solid waste materials formed. These collect on the inside of the cylinder in the form that we call carbon, and are eliminated when we take the car to a garage to have the carbon removed.

The above discussion concerns all kinds of internal combustion engines. Next we shall consider one particular type.

The Gas Engine

The gas engine is the type of internal combustion engine which today provides the motive power for all pleasure automobiles and most trucks. Beginning around 1680, attempts were made to devise an engine that would run on some more concentrated form of fuel than wood or coal. Among the fuels tried were gunpowder, turpentine, and illuminating gas. The first machine that operated on the same principle as a modern engine did not appear until 1878. The builder was a German named Otto. (In other words today's "auto" engine is a direct descendant of the original Otto engine.) It is a little startling to think that all the development from that first machine

to the high-powered automobile and airplane engines of today has occurred within a period of from 60–65 years.

Cycle of Operation. — Now we will trace the actual cycle of operation of the gas engine, following it through step by step for a one-cylinder unit. The relative positions of carburetor, intake manifold, and valve, piston and cylinder, spark plug, and exhaust valve and

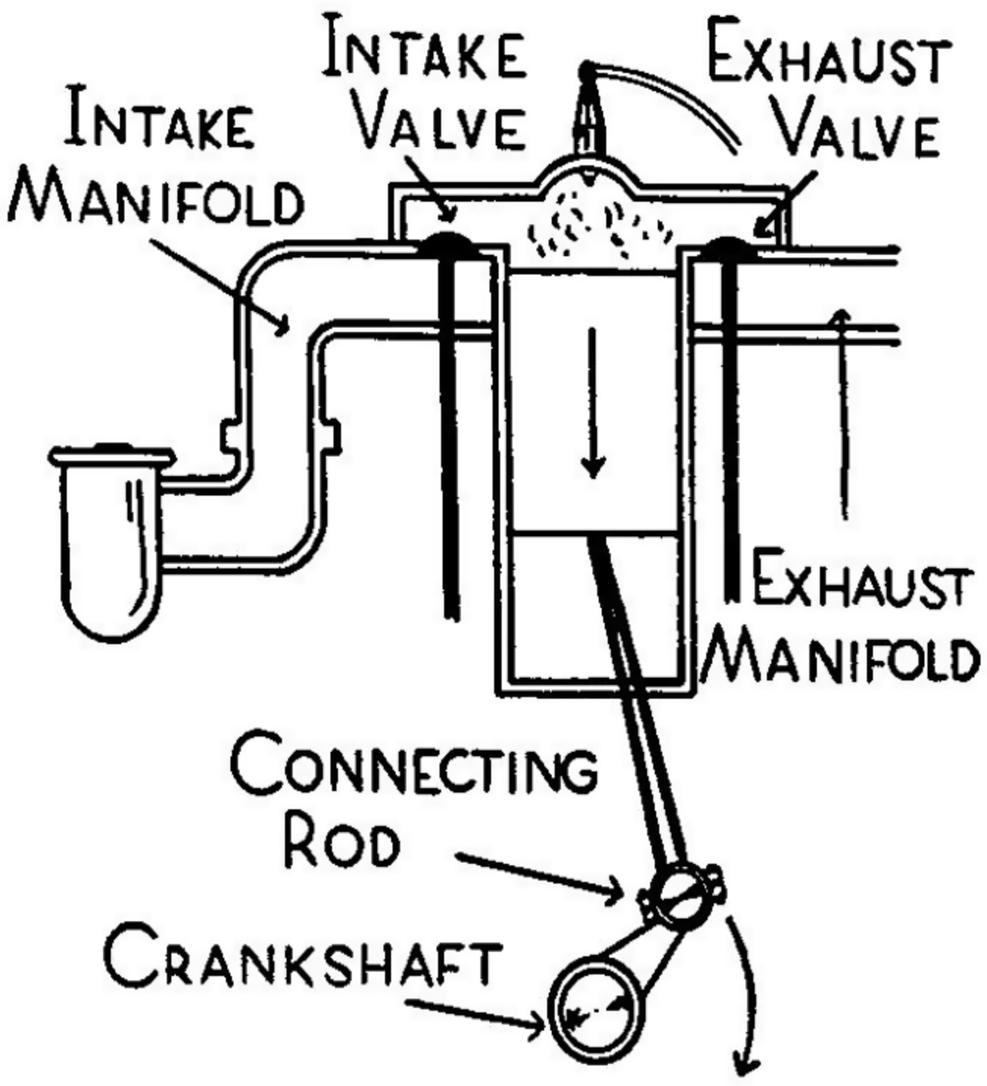


FIG. 114. Diagram showing the relative locations of various parts of a one-cylinder gas engine. (After a General Motors Corp. chart.)

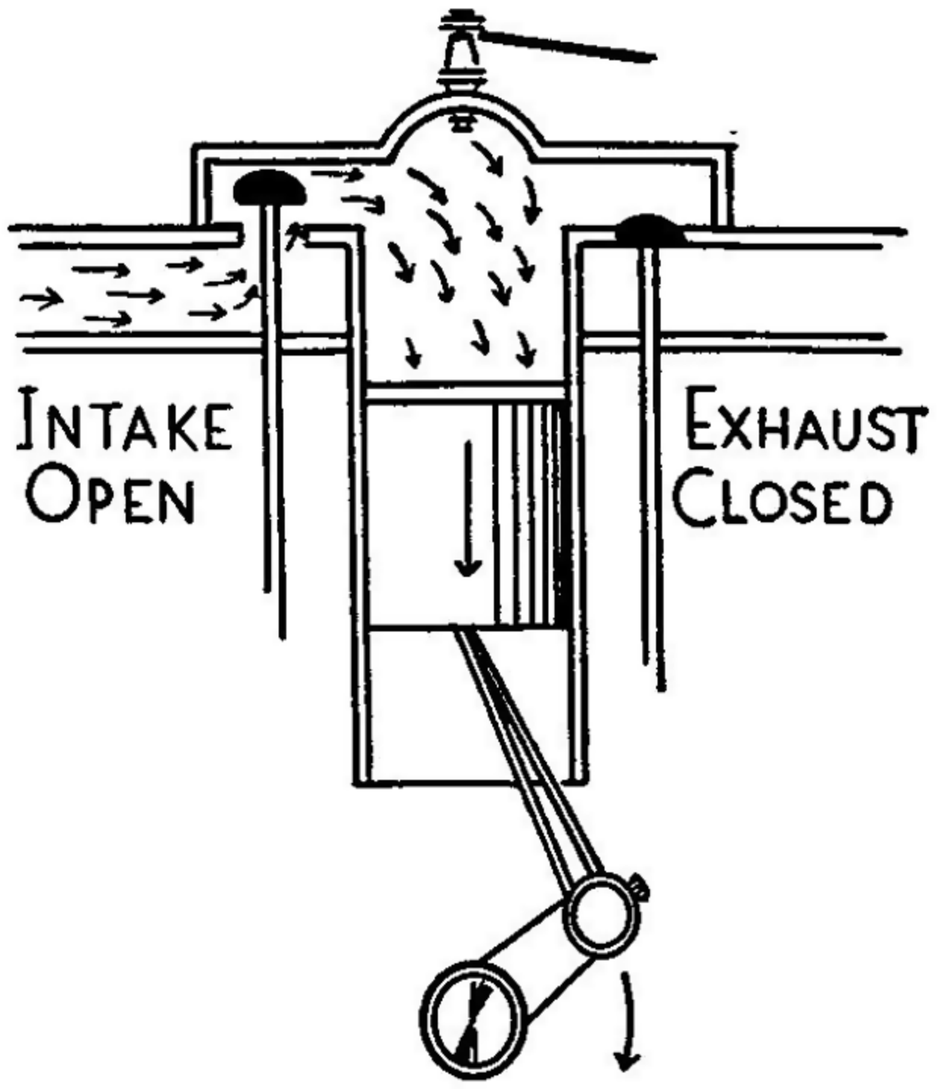


FIG. 115. Positions of piston and valves just after the beginning of the intake stroke. (After General Motors Corp. chart.)

manifold are shown in Fig. 114. Note also in this diagram the connecting rod running from the piston down to the crankshaft, the view of the latter being an end-on one. It should be recalled that here the term " stroke," as in the case of the steam engine, refers to a one-way trip of a piston along a cylinder.

In Fig. 115 piston, cylinder, valves, and so forth are shown a little after the beginning of what is called the *intake stroke*. As the piston moves down, a partial vacuum is created above it. Because of this the greater outside pressure forces the vaporized mixture of gasoline and air from the carburetor, along the intake manifold, through the open intake valve, and into the cylinder. The exhaust valve is closed during this stroke.

The next, or *compression stroke*, starts with the piston at the bottom of its path. Both valves are closed so that as the piston moves

upward, the fuel above it is compressed (Fig. 116). The extent of this compression is described by what engineers call " compression ratio." This may be defined as the ratio of the volume of gas in the cylinder when the piston is at its lowest point to the volume when the piston is at its highest point. If, at the beginning of the compression stroke, a certain cylinder contains a quart of vapor which, during the stroke, is squeezed down to one-fourth of a quart, the compression ratio is 4 to 1. Other things being equal, the higher the compression ratio the more efficient the engine; therefore, it is desirable to compress the gas as much as possible. On the other hand, compressing a gas raises its temperature, as anyone knows who has ever inflated an automobile tire or even a basketball or football with a hand pump. If the compression ratio is too high the heat generated will automatically ignite the fuel too quickly. Thus an upper limit is placed on practical compression ratios. The addition of anti-knock compounds to gasoline has made it possible to use compression ratios of around 6 to 1 in modern automobile engines. These substances, usually compounds of lead, serve the purpose of keeping the gasoline vapor from igniting quite so readily as it otherwise would.

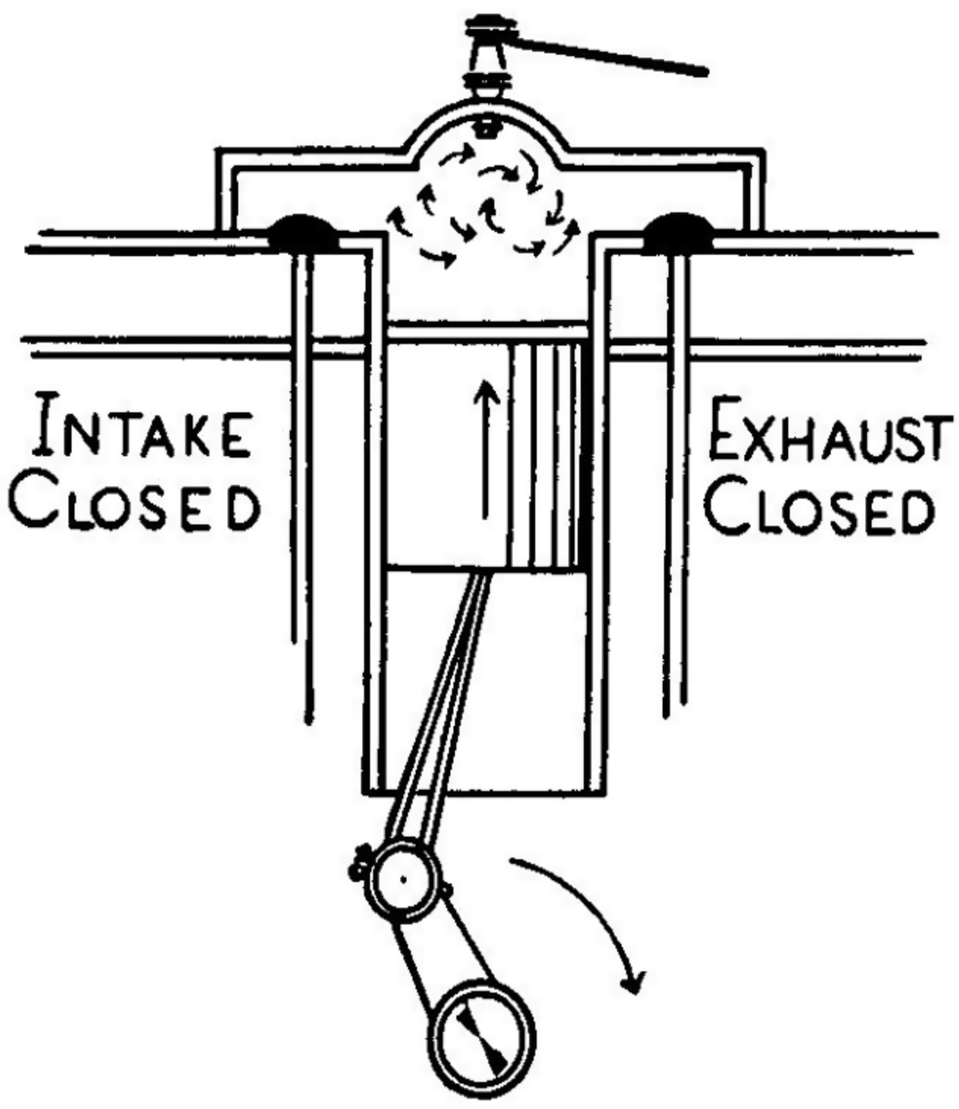


FIG. 116. Positions of piston and valves toward close of compression stroke. (After General Motors Corp. chart.)

Shortly before the piston reaches the end of the compression stroke and starts down again, an electric spark jumps across the gap at the bottom of the spark plug, igniting the fuel. By the time the burning occurs, the piston has just passed its highest point and started down again. The force of expanding gases drives the piston down and is the motive power for the operation of the engine (Fig. 117). This is called the *power stroke*. Obviously, both valves must be closed throughout this stroke also, since otherwise the explosive force would not all be directed against the piston. It is principally in connection with the compression and power strokes that the presence of carbon

in the cylinder reduces the effective operation of the engine. These solid materials collect around the valves, prevent them from closing tightly, and so cause loss of compression and power. Carbon in the cylinder also prevents the escape of heat which increases the chances of automatic ignition of the gas.

During the last, or *exhaust stroke*, of the cycle, the piston travels upward again (Fig. 118). In doing this it forces the waste products

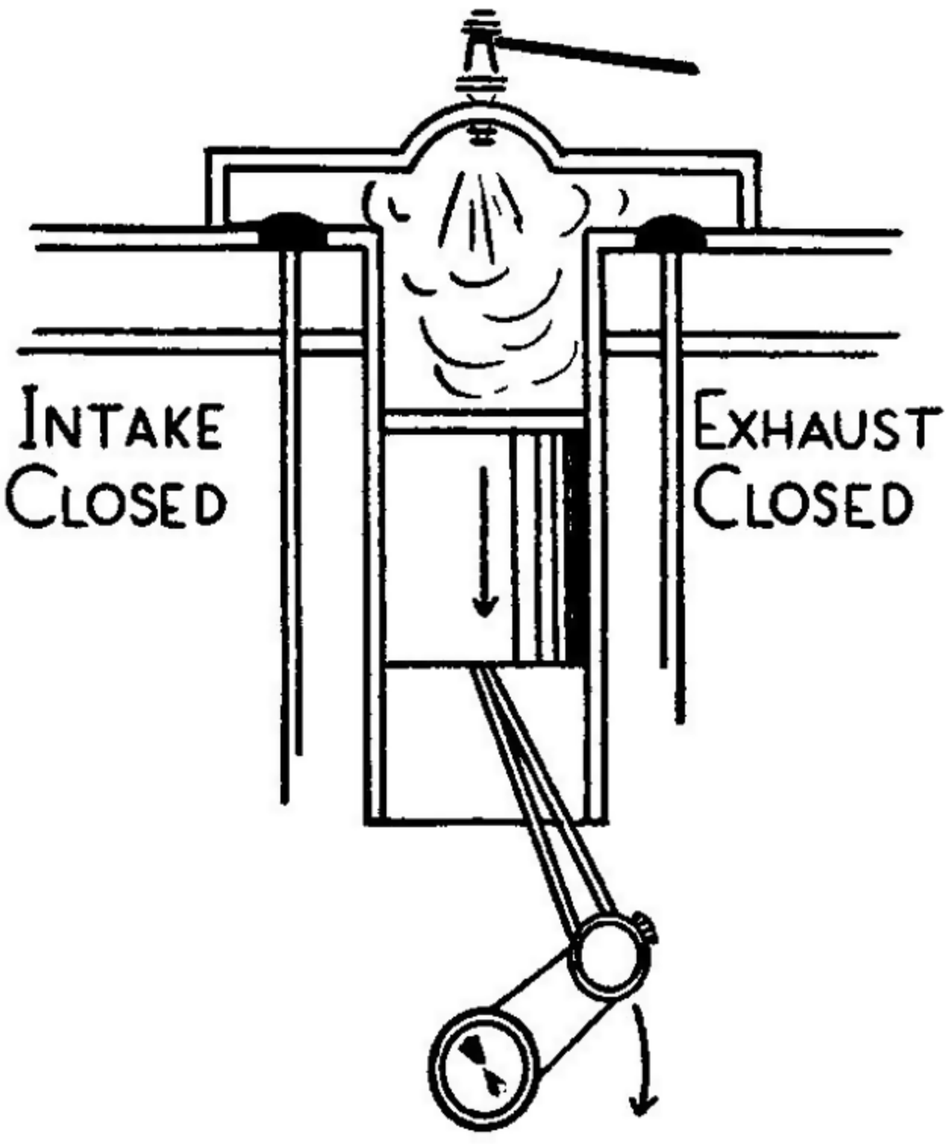


FIG. 117. Positions of pistons and valves shortly after beginning of the power stroke. (After General Motors Corp. chart.)

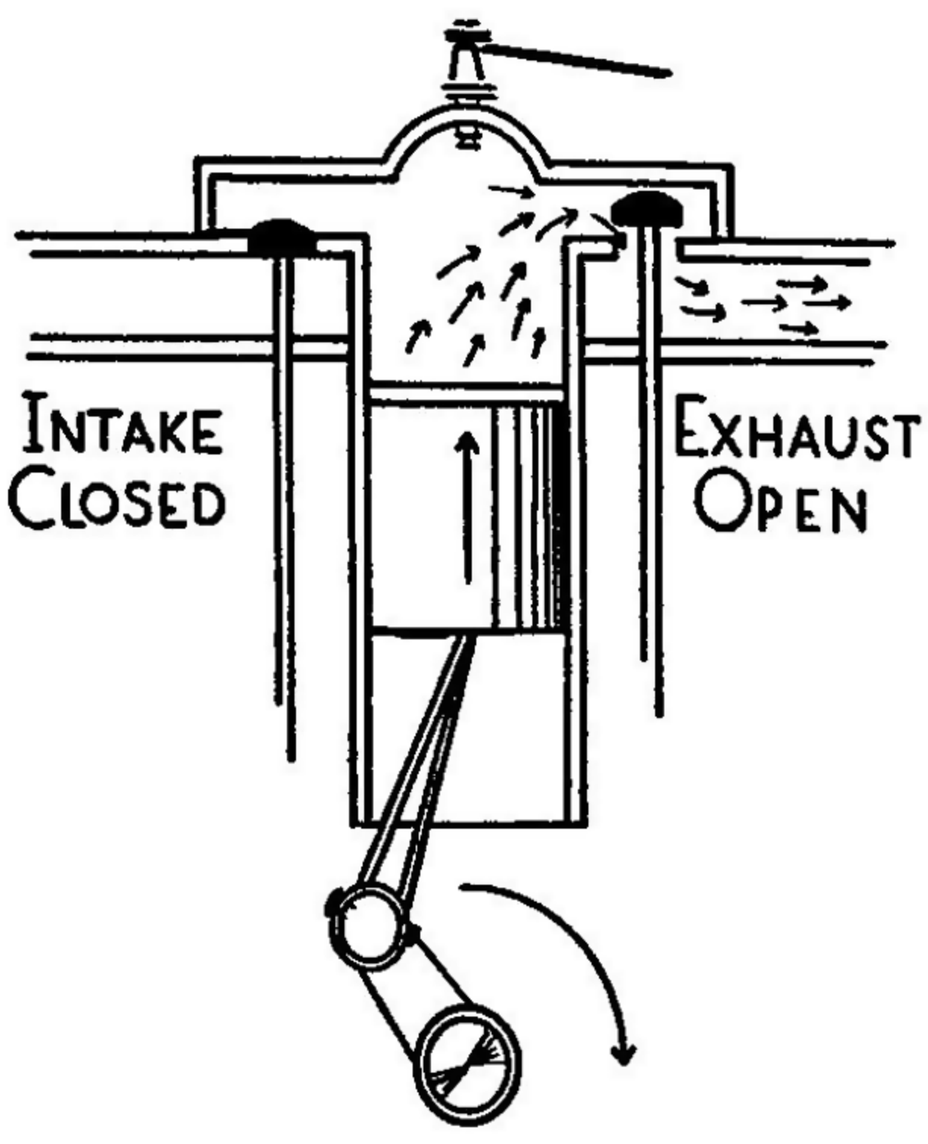


FIG. 118. Positions of piston and valves toward close of exhaust stroke. (After General Motors Corp. chart.)

of combustion out through the open exhaust valve into the exhaust manifold and thence to the exhaust pipe. The intake valve is closed since it would not be desirable to drive part of the waste gases back into the carburetor.

Thus one cycle is completed. The four parts of it are repeated in sequence as long as the engine runs.

It will be noted that during one cycle the piston makes four strokes. Any gas engine, regardless of the number of cylinders it possesses, that operates on this cycle is said to be a " four-stroke-cycle engine," a name that is frequently but erroneously abbreviated to " four-cycle engine." It should also be pointed out that since the crankshaft makes a half revolution during each stroke, it makes two complete revolutions for every cycle.

The operation of the valves is taken care of automatically by a cam and a camshaft which is geared to the crankshaft. A device called a timer acts as a sort of time-clock and sends a surge of electrical energy through the ignition circuit at exactly the instant that it is desired to have the fuel ignited in the cylinder.

There is also a two-stroke-cycle gas engine in which, as the name implies, an explosion occurs every second stroke. It is of considerably less importance than the four-stroke-cycle type and will not be discussed here. Readers who are interested in the details of its operation should consult a textbook in physics or internal combustion engines, or an encyclopedia.

Multiple-Cylinder Engines. — In the one-cylinder engine just discussed, only one of the four strokes furnishes the power that is required in order for the engine to coast through the other three stages of the cycle. Actually, power is being exerted on the piston less than one-fourth of the time, since the force resulting from combustion is almost entirely dissipated before the piston reaches its lowest point. The motion of such an engine is not smooth but consists rather of a series of surges separated by the longer coasting periods. A heavy fly-wheel serves to smooth out to some extent the motion of a one-cylinder gas engine by making use of our former acquaintance, inertia. Both smoother operation and increased power are obtained by using engines of more than one cylinder.

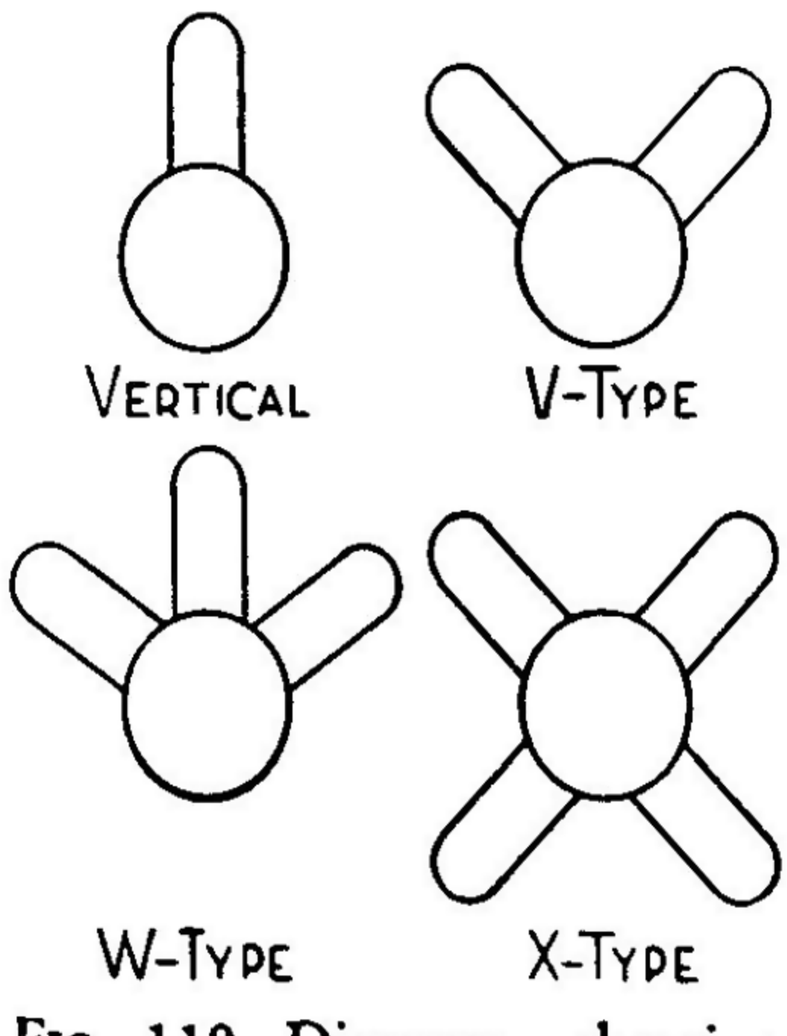


FIG. 119. Diagram showing end-on views of four types of liquid-cooled airplane engines.

Consider a four-cylinder gas engine. In the first place, the crankshaft will have to be long enough to have four throws or connecting rod connections in it. Since there are four strokes in the cycle and four cylinders, the timing can be so arranged that some one of the cylinders will fire every half revolution of the crankshaft. However, because of the fact mentioned above, that force is not exerted through the entire power stroke, even in a four-cylinder engine there are times when no power is being applied to the crankshaft. In a six-cylinder

engine the power strokes overlap slightly. With more than six cylinders the overlapping, of course, increases.

On the other hand, multiplying the number of cylinders has the disadvantage of increasing the cost of operation, other things being equal.

Gasoline engines used in propeller driven airplanes are of the

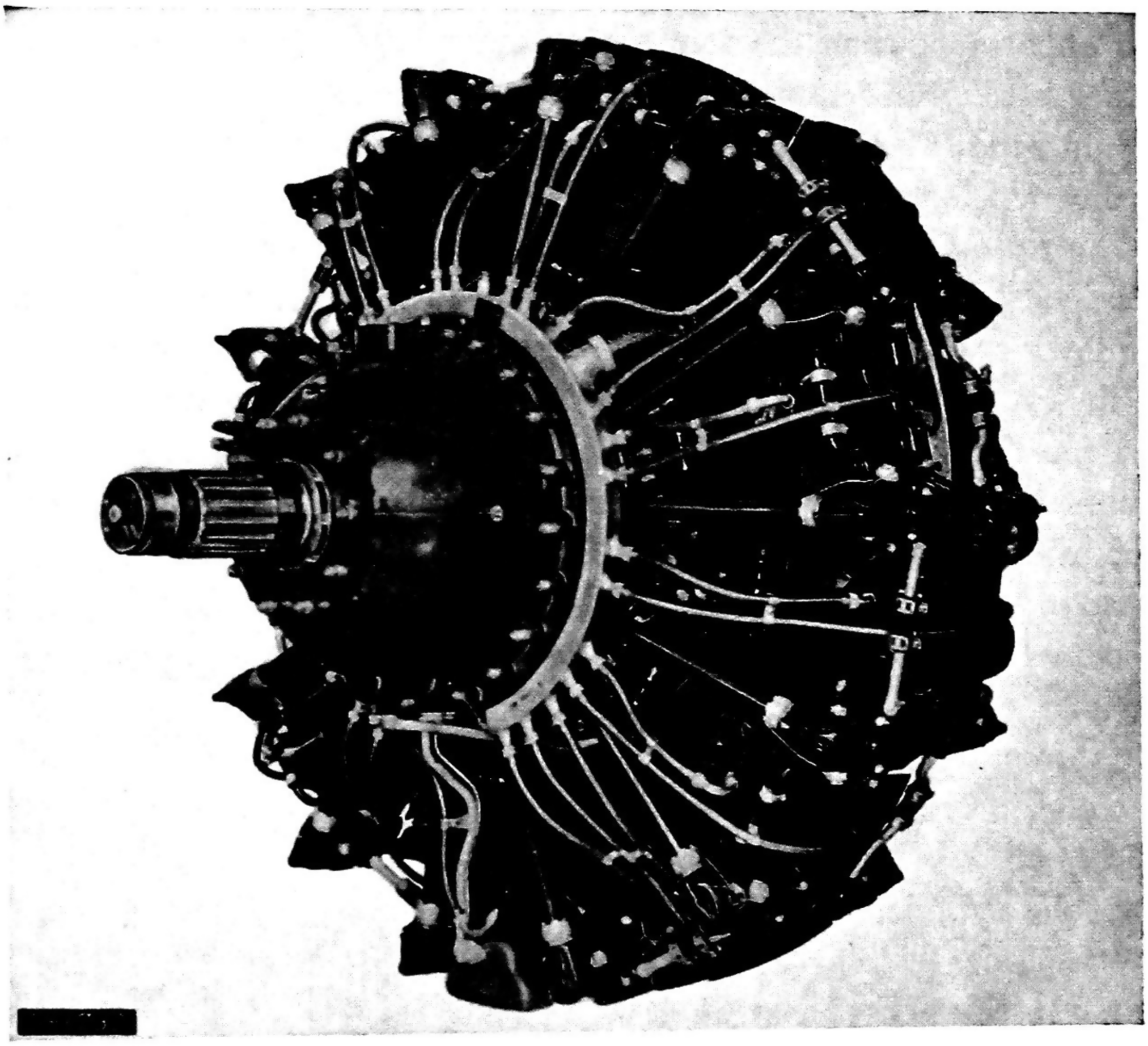


FIG. 120. A 1200-horsepower Pratt & Whitney R-1830 air-cooled airplane engine. (United Aircraft Company.)

four-stroke-cycle type and differ little from automobile engines in the essentials of their construction and operation. Among the liquid-cooled airplane engines there are, in addition to the vertical- and V-types used in automobiles, so-called W- and X-type engines. Comparative diagrammatic end-on views of these engine types are shown in Fig. 119. The air-cooled or radial type airplane engine, in which the cylinders are grouped around a crankshaft like the spokes of a

wheel, also operates on a four-stroke cycle. It differs from the other type in that the crankshaft has only one " throw " in it. One master connecting rod is attached to this throw and the other rods to the master rod. Such engines are built with odd numbers of cylinders in one bank or row. When the engine has two banks of cylinders there are two throws in the shaft.

Gas Engine Efficiencies. — The average efficiency of gas engines as a whole, from fuel through to useful work, is about 25 per cent, with the best ones reaching something like 30 per cent, which is appreciably better than the average for steam engines.

Data from the Research Division of the General Motors Corporation indicate that the over-all efficiency of the average automobile is about 8 per cent, which means that only 8 cents of every dollar spent for gasoline is expended in actually moving the car along the road. The remainder is lost in various ways chief among which are 40 per cent absorbed by the water in the radiator and 20 per cent unavailable because of incomplete combustion. The principal hope for increasing the total efficiency lies in reducing these combustion losses by improving techniques of fuel injection, ignition and burning.

The useful work in aircraft may run as high as 25 per cent, the difference being due principally to the low frictional and heat losses in comparison with those mentioned above.

A comparison of different engine efficiencies is shown in Fig. 121.

The Diesel Engine

The Diesel engine, which has excited a great deal of popular interest in recent years, is inherently more efficient than either the steam or gas engine. The reason for this will be pointed out in the course of the discussion.

Comparison with Gas Engines. — In outward appearance a modern Diesel engine bears a very close resemblance to a gas engine, except that for equal-powered units, the former is somewhat the larger of the two. Both are heat engines. Both machines use a liquid fuel which is transformed into vapor form before being ignited. Both have cylinders in which pistons move with reciprocating motion. Both have intake and exhaust manifolds and valves. Both transform

the reciprocating motion of the pistons into rotational motion of a shaft by means of connecting rods and a crankshaft. Both operate, in most installations, on a four-stroke cycle, although there are also two-stroke-cycle types of both engines. The four strokes of the cycle are given the same names in the two machines.

There are three principal respects in which they differ. One is in the fuel employed. Whereas the gas engine of today commonly

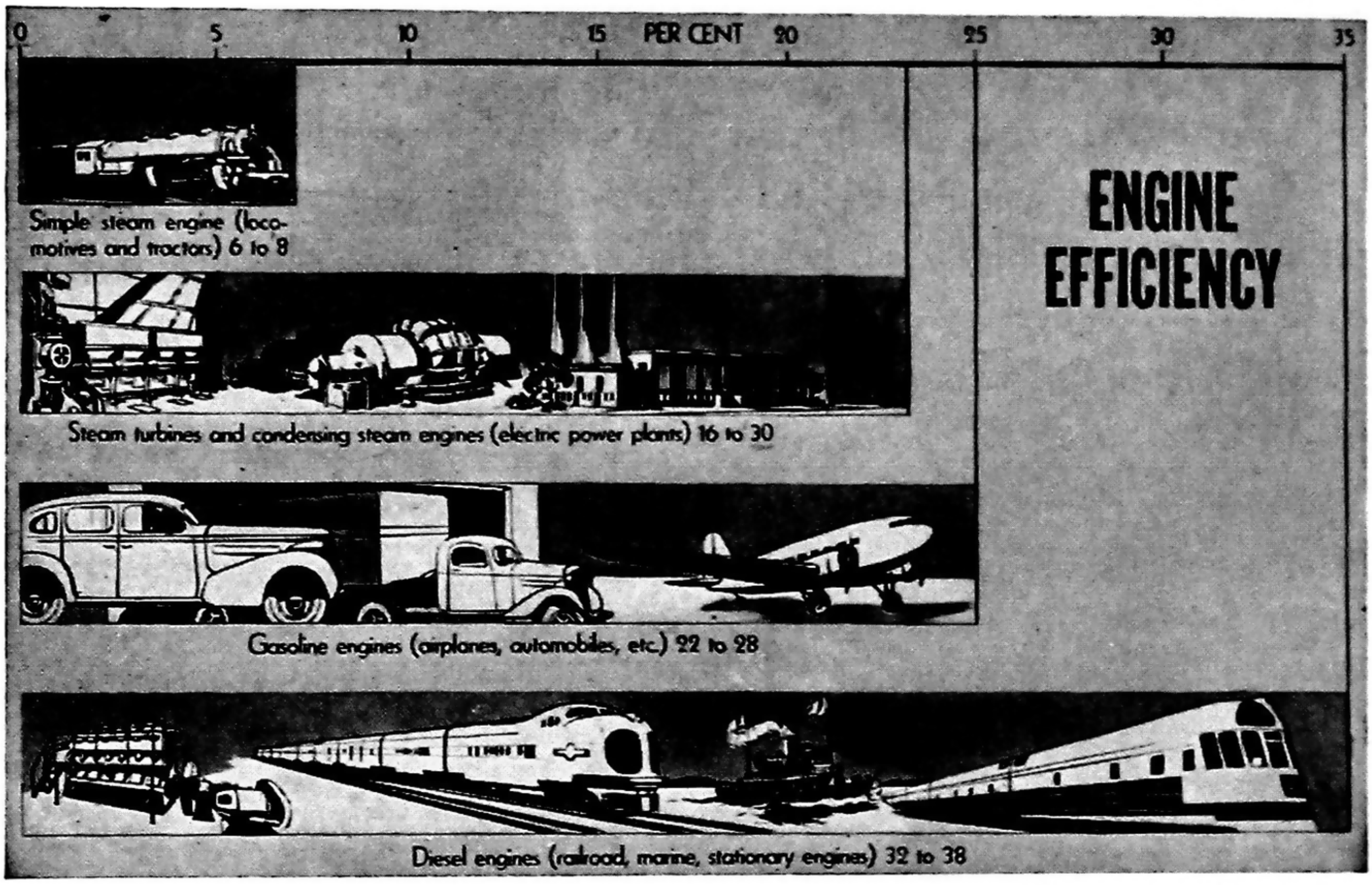


FIG. 121. Diagrams showing efficiencies of various types of engines. (General Motors Corporation.)

uses gasoline, Diesels in general are run on a rather highly refined type of fuel oil. The relation of this fuel situation to the comparative costs of operation of the two engines will be considered a little later in this chapter. The other two important differences concern compression and ignition of the fuel and can best be described in connection with those strokes of the cycle.

Cycle of Operation. — The *intake stroke* in the Diesel engine differs from its counterpart in the gas engine in that only air is drawn into the cylinder instead of a mixture of air and fuel.

Since nothing but air enters the cylinder during the intake stroke, it is obvious that only air can be compressed during the *compression stroke*. The reason for this variation from the gas engine case con-

cerns the matter of the heat developed by compression. Whereas maximum compression ratios in gas engines are in the neighborhood of 6 to 1, they may run as high as 16 to 1 in Diesels. This enormously greater compression creates temperatures as high as 1,000° F. which is sufficient to ignite automatically any fuel that is present. If prema-

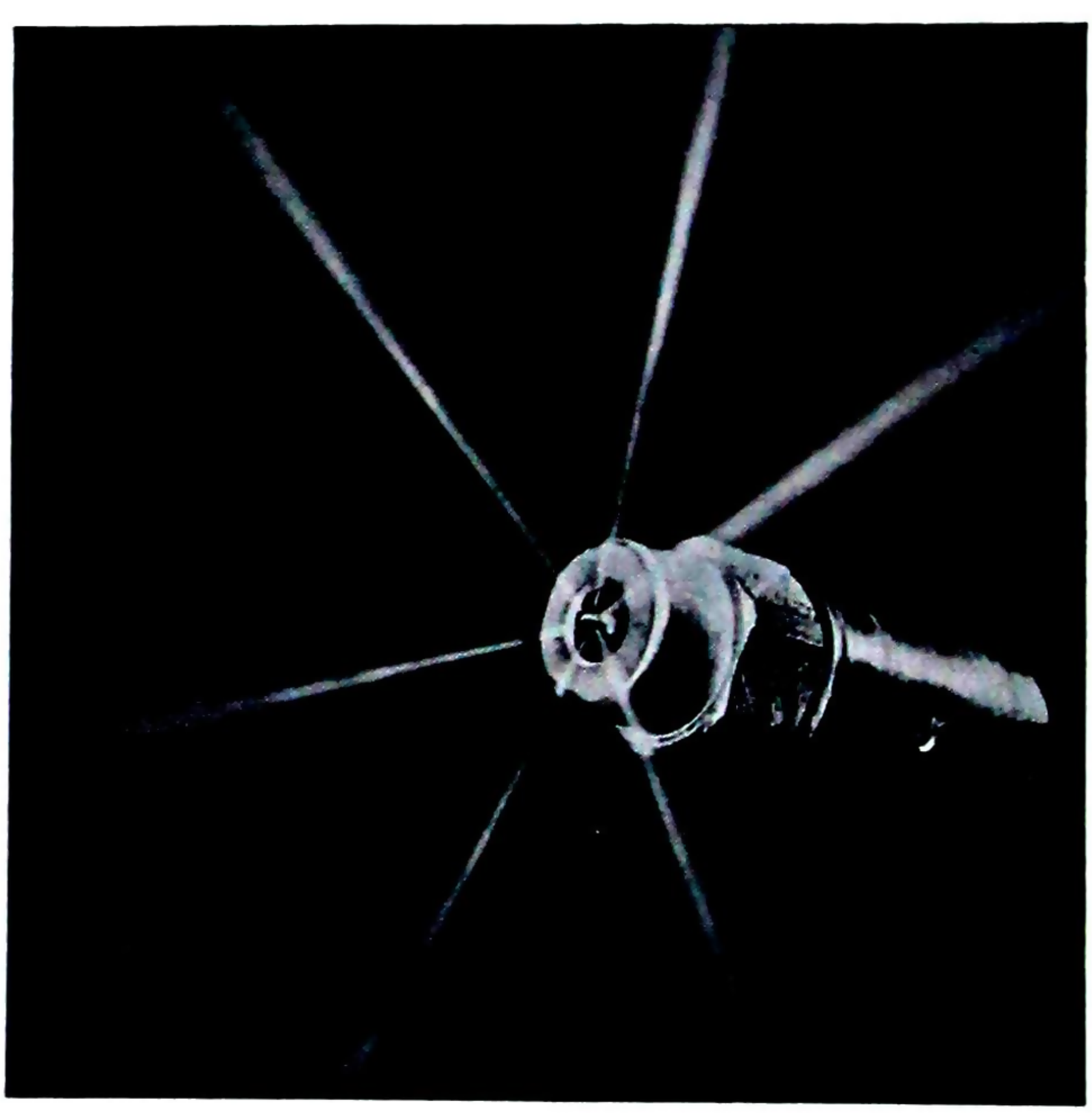


FIG. 122. Photograph of the oil spray coming from a Diesel fuel injector. (General Motors Corporation.)

ture ignition is to be avoided, the cylinder must not contain any fuel during compression.

In the *power stroke*, ignition of the fuel occurs at the instant it enters the cylinder, because of the high temperature generated by compression. The fuel is sprayed into the cylinder by an injector device. Figure 122 is a photograph of such an oil spray as it comes from the tip of one type of fuel injector. Such devices force the oil into the cylinder at pressures of from 3,000 to 20,000 pounds per square inch, and at velocities as high as 13 miles per minute.

The action in a Diesel engine during the *exhaust stroke* is exactly the same as that described for a gas engine.

Advantages and Disadvantages. — Since the gas engine is the Diesel's principal competitor, one tends to classify its features as good or bad in comparison with the other type of machine.

The principal advantage, as was suggested at the beginning of the discussion of Diesels, is its greater economy of operation. This is true because of the experimental fact that, other things being equal, engine efficiency increases with the compression ratio. It is true that modern Diesels use approximately half the quantity of fuel consumed by gas engines for the same horsepower unit.

One popular misconception about Diesel engines should be mentioned in this connection. That is the common belief that Diesels will operate satisfactorily on very poor fuels such as old crankcase oil. While it is true that Diesel engines have been run on various things including powdered coal, potato peelings, sawdust, pine needles, hay, rice hulls, thin tar, buttermilk and a few others, it is also true that they run well and efficiently only on a rather highly refined type of fuel oil that costs about the same as gasoline at the refinery. So their greater economy of operation is not due fundamentally to their being run on cheap fuel.

Among the disadvantages are the facts that the first cost of Diesels is somewhat greater than that of gas engines, and that the weight per horsepower is also higher. It is also true that the Diesel engine is the less flexible of the two and does not run well over as wide a range of speeds.

Uses. — Since Diesel engines are now built in sizes from 5 horsepower up to more than 20,000 horsepower, their applications are quite varied. In stationary installations they are used to generate power in cases all the way from small factories up to whole towns. More than half of the world's shipping today is carried in Diesel-powered boats. Other applications include their use in trucks, tractors, road-building equipment, and streamline trains.

The Reaction Engine

The general term " reaction engine " applies to a group of power plants which have received an increasing amount of attention in recent years — particularly for their military applications. Names applied commonly, though frequently loosely, to various members of

the group include rocket, jet-propulsion engine, thermal-jet engine, duct engine and others. A little later we will define these types and differentiate among them but first the group as a whole will be considered briefly. The distinguishing characteristic of all reaction engines is their common principle of operation which is Newton's third law of motion. This law, it will be recalled, states that *to every action there is an equal and opposite reaction.* For example, a man jumping from a rowboat exerts a backward thrust on the boat, pushing it in the opposite direction; a bullet fired from a gun causes the gun to "kick"; water emerging from the nozzles of a rotary lawn sprinkler exerts a backward thrust which produces the rotation. Generation of motion in a reaction engine is just that simple as far as principle is concerned; a gas emerging from a jet exerts a thrust in the opposite direction against the chamber from which it issues causing the engine, and the airplane or other device to which it is attached, to move.

The idea of producing motion in this manner is not new. In the first or second century A.D. the steam reaction engine, or aeolipyle, shown in Fig. 123 was invented by Hero of Alexandria, a Greek mathematician and scientist. Steam generated in the hollow sphere issued from the side spouts and the reaction or "kick-back" caused the outfit to rotate. This particular reaction engine, however, was never put to any practical use.

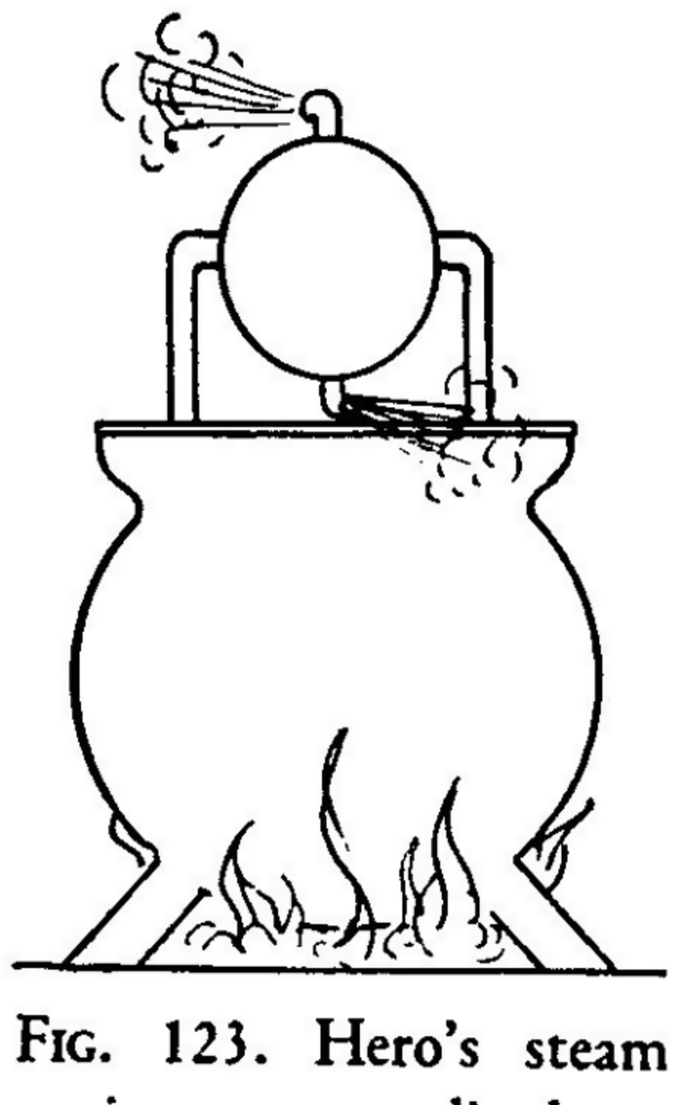


FIG. 123. Hero's steam engine, or aeolipyle — about 150 A.D.

In the reaction engines which we will consider, a fuel of some sort is burned and the escape of the resulting gaseous products of combustion through a jet in the rear of the device causes the desired reaction or thrust forward. Two major types of reaction engine can be distinguished on the basis of the source of the oxygen used for combustion. (The role of oxygen in the burning process is discussed in an earlier chapter.) These are airstream engines, which use atmospheric oxygen to burn the fuel they carry, and rockets, which contain both fuel and an oxidizing agent and are, therefore, independent of

the oxygen content of the air through which they travel. Reaction engines are heat engines in that, like the types discussed earlier in this chapter, they convert heat energy into mechanical energy. In fact, strictly speaking they probably should be classed as internal combustion engines since burning occurs within a combustion chamber. We are considering them separately, however, since by usage the term "internal combustion" ordinarily signifies the type of engine in which a piston is driven back and forth in a cylinder by the forces resulting from a series of explosions of a mixture of air and a vaporized fuel.

Airstream Engines

Airstream engines resemble the gas and Diesel types described previously in the broad features of their operation. Air is introduced into a combustion chamber; compression occurs; the air-fuel mixture is burned; and the resulting force is employed to produce motion. They differ in the particular manner in which these effects are accomplished with the airstream devices requiring few or no mechanically moving parts. These differences will be evident from the following discussion of three kinds of airstream engines.[1] Schematic diagrams of the three types are given in Fig. 124.

The Turbo-Jet Engine. — The operation of all turbo-jet (also called thermal-jet) engines can be described in terms of the top engine of Fig. 124. This figure does not depict any specific design but indicates the general makeup of all airstream engines in which compression is effected by a turbine-powered compressor. Air, drawn in at the front, passes through the vanes of the rotary air compressor and reaches the combustion chamber at high pressure. Fuel is injected into the chamber through the fuel nozzles and burns in the compressed air. The greatly accelerated gases are then ejected from the jet at the rear, imparting an opposite thrust to the engine in the direction shown by the large arrow. Power to operate the rotary compressor is supplied by a turbine wheel located at the rear of the combustion chamber (see diagram) and attached to the compressor shaft. The turbine wheel is driven by the escaping gases in the same

[1] The airstream engine sub-groups listed here are those used in *The Coming Age of Rocket Power*, by Edward Pendray, Harper & Brothers, 1945.

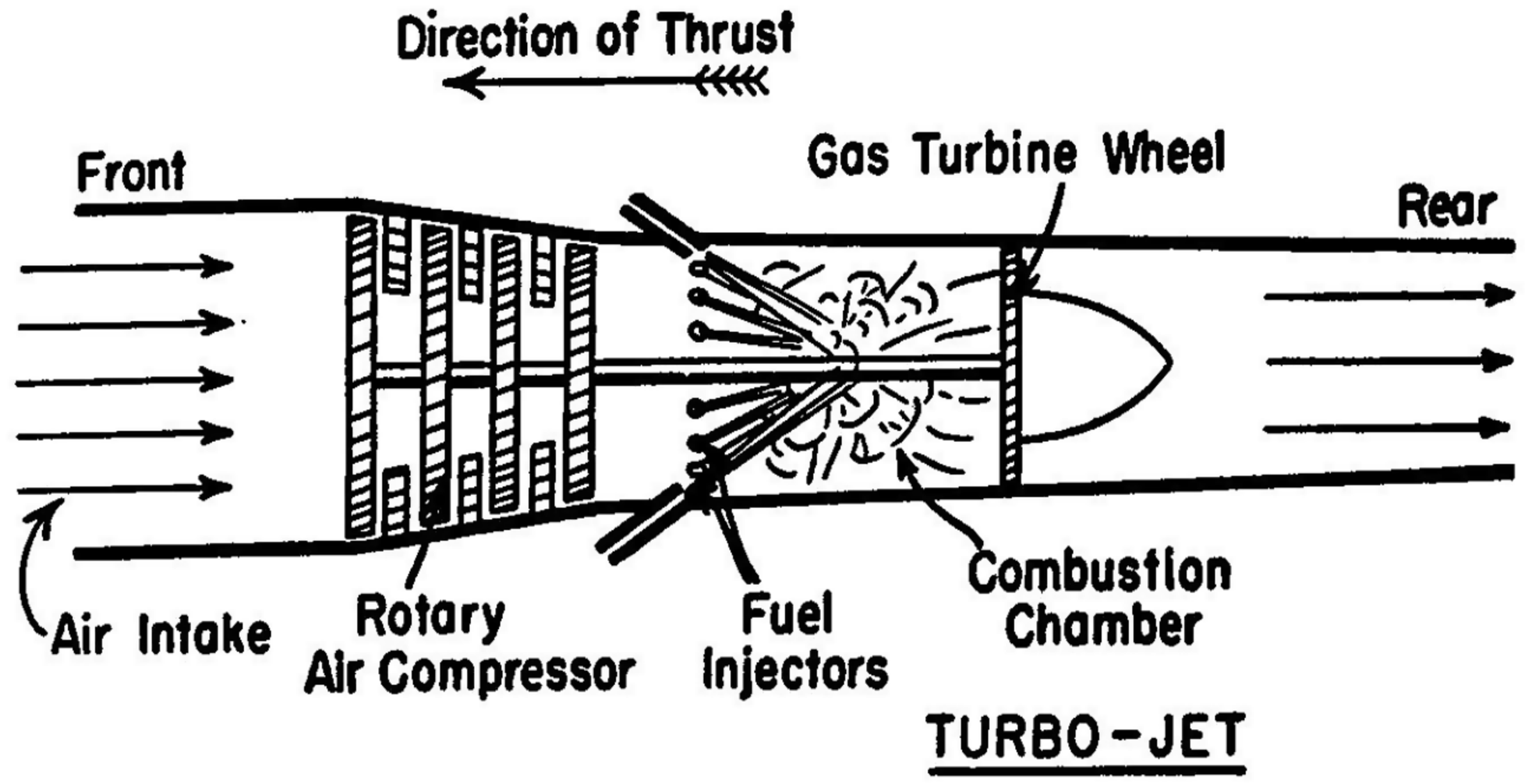


FIG. 124. Three types of airstream engines: (a) turbo-jet, (b) intermittent duct, (c) continuous duct. (After G. Edward Pendray's *The Coming Age of Rocket Power*, Harper & Brothers.)

manner that wind drives a windmill or steam a simple steam turbine. Since the compressor must be turning before the engine can operate, either compressed air or an electric motor is usually used to start the unit. Initial ignition of the fuel can be accomplished by an electric spark. Once started, operation is continuous as long as fuel is fed into the chamber. Fuel requirements, as far as the engine is concerned, are simply that the substance burned have a high heating value and be easy to introduce into the combustion chamber. These are easily met by existing petroleum products. Certain other restrictions may arise, however, as the engine finds special applications as, for example, in airplanes operating at extremely high altitudes.

Among the principal advantages of the turbo-jet engine is its low weight. It has been predicted that future gas turbines will not weigh more than half as much per horsepower developed as does the best conventional internal combustion engine. One of its big disadvantages at present is its high fuel consumption.

This type of power plant, which has but one moving part in the engine itself, involves no gears, pistons, cams, connecting rods or crankshafts, and applies its thrust directly instead of through the medium of wheels or a propeller, is becoming increasingly important, particularly in the field of aviation. Figure 125 shows the U. S. Army Air Force's new turbo-jet-propelled P-80 fighter, the "Shooting Star," which has attained speeds in excess of 550 miles per hour. Its service ceiling is above 40,000 feet. In this plane air for the jet engine enters by vents in front of the wings close to the fuselage (see figure). The fuel used is aviation kerosene. The jet exhaust is at the rear tip of the fuselage. There is virtually no vibration, which reduces pilot fatigue. Sufficient power for takeoff can be developed in less than a minute.

The Intermittent Duct Engine. — Both the intermittent and continuous duct engines differ from the turbo-jet unit in that they contain no compressor; in fact they contain almost no moving parts whatever. Air which enters the front vents at high speed exerts a ram action on the air already in the chamber compressing it. One example of an intermittent duct engine—a German model patented in 1930—is shown in the center picture of Fig. 124. Its cycle of operation is as follows: First, the engine must be moving rapidly

forward (150 miles per hour or more) before it can function. As a result of this motion air enters through the air inlet valves which are designed to open easily to let air in but to close if air begins to flow out through them. As the entering air compresses that already in the chamber, fuel is injected and burns. The resulting explosion closes the valves and forces out the rear nozzle a blast of gases which

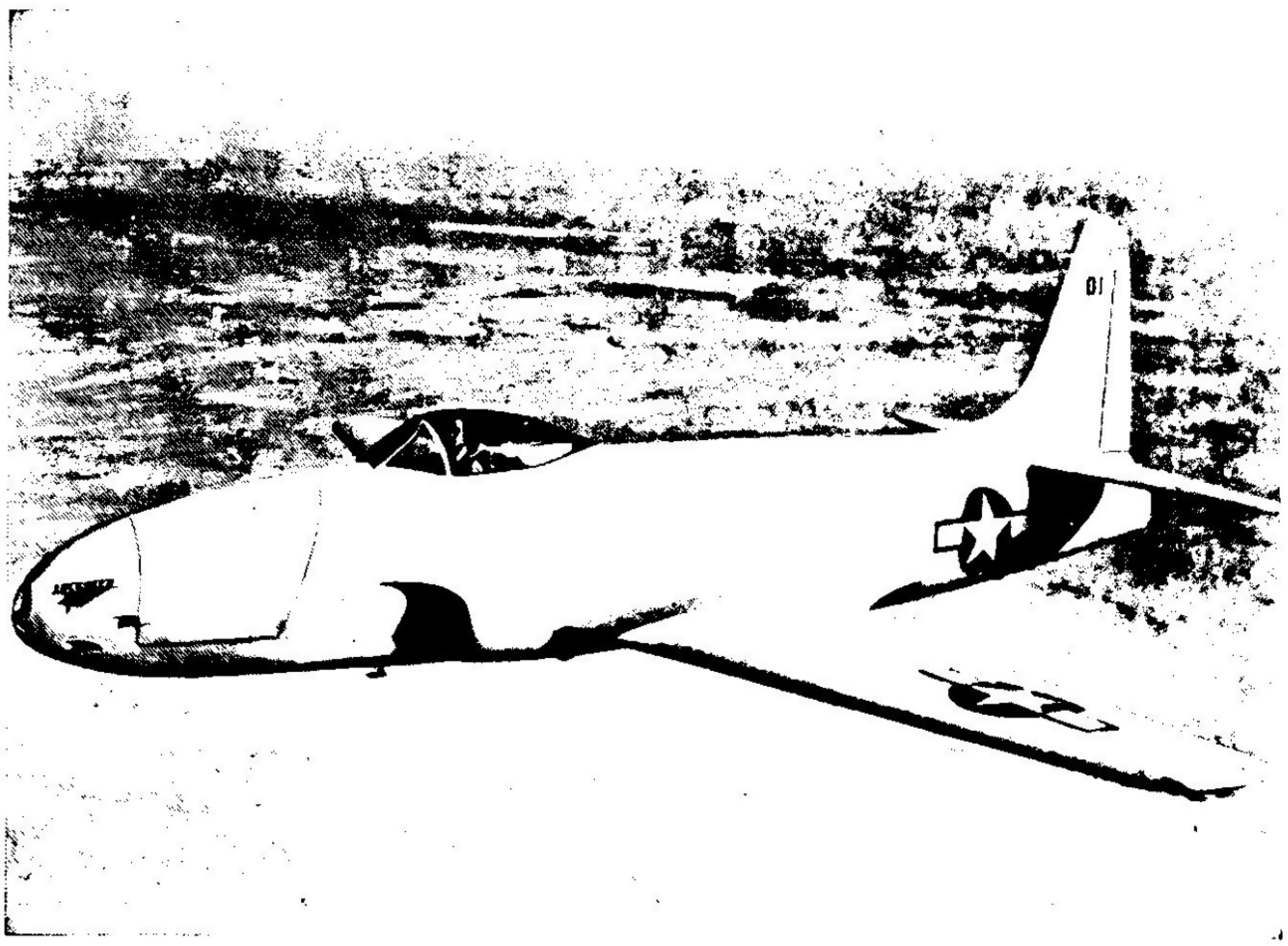


FIG. 125. U. S. Army Air Force jet-propelled P-80 "Shooting Star." (U. S. A. A. F. Photo.)

exerts a momentary opposite thrust in the direction shown. Immediately following the explosion there is a low pressure region just inside the valves which permits the force of the outside air to open them and begin the cycle over again. Generation of thrust is, of course, pulsating or intermittent.

Although this type of engine has not been employed to power full-sized airplanes it was used in the famous German V-1 robot bomb, or "buzz-bomb." The engine was mounted above and at the rear of the flying bomb which was airplane-shaped and had a length of some 25 feet and a wingspread of about 16 feet. The duct engine was approximately 11 feet long. In the large forward end of the engine was a lattice containing several hundred square holes through

which the air entered. Behind these holes were thin steel flutter valves. Fuel was injected by means of compressed air through nine jets. A spark plug was used for initial ignition; after the first few explosions enough flame remained after each one to provide ignition for the next. The explosions occurred at a rate of approximately 2,400 per minute. The speed of the bomb was between 250 and 400 miles per hour. Its rate of fuel consumption was some eight times that of an internal combustion engine of equal power.

The Continuous Duct Engine. — This engine, which also goes by the names of ram-jet, athodyd, and flying stovepipe, is the simplest of the three airstream power plants. It consists simply of a tube open at both ends and shaped as indicated in the bottom diagram of Fig. 124 with provision for fuel injection and ignition. Before it will function the engine must be moving through the air at high speed and, in general, operates most effectively at speeds greater than the speed of sound. Air entering the front slows down momentarily as a result of the increased diameter of the tube. Because it slows down it is compressed by the ram effect of the inrushing air following it. Combustion occurs in the mixture of fuel and compressed air just as in the case of the other engines. The gaseous products of combustion are driven out the rear and exert a reaction which provides a continuous net thrust in the direction shown.

No applications of this power plant to airplanes have yet been tried due to the high speed required to make it operate. However, as part of the U. S. Navy's guided-missile program, the Applied Physics Laboratory of Johns Hopkins University had, by mid-1946, successfully flown 70-pound ram-jets, developing over 2,000 horsepower each, at speeds greater than 1,400 miles per hour.

Rockets

Rockets which constitute perhaps the simplest form of reaction engine can be divided into solid-fuel and liquid-fuel types. Both kinds carry an oxidizing agent as well as fuel; thus combustion of the fuel, and consequently flight, does not require the presence of oxygen in the atmosphere through which the rocket travels. This means that, whatever the other difficulties may be, at least lack of

oxygen in the space between the earth's atmosphere and the moon would not preclude a lunar trip by rocket. Airstream engines, on the other hand, cannot operate in an atmosphere devoid of oxygen. If the reader is concerned because the jet of a rocket traveling in a vacuum would have no air to push against he should recall again Newton's third law. A jet-powered device (projectile or airplane) moves, not because the jet pushes against the air, but because it reacts or " kicks " back against the engine.

Solid-Fuel Rockets. — The common or garden variety Fourth of July skyrocket is the best known engine of this type. The propellant, which is usually black powder mixed with something to slow down burning, is packed into a heavy paper tube. The walls of the tube at the rear of the charge are shaped to form a nozzle while the rear surface of the fuel itself is the shape of a hollow cone. A fuse runs from the apex of this cone out through the nozzle. The hollow cone forms the combustion chamber as burning starts and the reaction to the rush of gases out the nozzle starts the rocket on its way. As burning progresses, of course, the combustion chamber enlarges.

Among the solid-fuel rocket's chief disadvantages are the loss of combustion control once the fuel is ignited and the rapid decrease in mass that occurs as the fuel burns. The latter characteristic changes the stability of the device as it travels through the air.

Military applications of solid-fuel rockets in the last war included the bazooka projectile, and airplane and antiaircraft rockets.

The Liquid-Fuel Rocket. — Use of solid-fuel rockets is confined to short-range applications. For longer flights—such as long-range military rockets—liquid-fuel types are required. The principle of operation is essentially the same; combustion, however, instead of occurring in a cavity in the fuel takes place in a metal chamber into which liquid fuel and oxygen are injected. One of the principal problems is the development of metal alloys able to withstand the high temperature of burning. Better control of combustion is possible but rate of fuel consumption is very high.

The most spectacular war application of the liquid rocket was in the German V-2 bomb which traveled at supersonic speed. These missiles, which were 40 to 50 feet long, burned alcohol and liquid oxygen. The explosive load was about one ton and the total weight

of the bomb at launching approximately 12 tons, two-thirds of which was fuel. Fuel consumption was at the rate of some 200 pounds per second which meant that the V-2 used about seven tons of fuel in its approximately 20 seconds of powered flight. At the mid-point of their flights these bombs reached maximum altitudes of from 65 to 70 miles; the horizontal distance covered was some 200 to 250 miles.

Questions

Discussion

1. Name ten specific examples of the use of energy that can be traced back to a fuel of some sort.

2. The most efficient steam engines and turbines produce approximately how many foot-pounds of useful work per B.t.u. of heat energy in the fuel?

3. Mention some advantages possessed by the Newcomen engine over the Savery engine.

4. Calculate exactly what height of water would be required to exert a pressure of 15 pounds per square inch.

5. Name the improvements made by Watt on the Newcomen type of engine and explain wherein each one really was an improvement.

6. Explain the advantages of a condensing steam engine over a non-condensing one.

7. Compare and contrast the steam engine and gas engine with respect to intake of fuel, ignition of fuel, and exhaust of waste products.

8. Discuss the function of each of the following parts of an automobile engine: (a) carburetor, (b) spark plug, (c) crankshaft, (d) timer.

9. Draw diagrams showing the location of the piston in the cylinder and the position of the valves at the *close* of each of the four strokes of a gas-engine cycle.

10. Explain why there is an upper limit to the practical compression ratios that can be used in automobile engines.

11. Connect the earlier discussion of inertia with the statement in this chapter to the effect that a heavy fly-wheel helps smooth out the motion of a one-cylinder gas engine.

12. Find out why there are 3-, 5-, 7-, and 9-cylinder radial aircraft engines, but not 2-, 4-, 6-, and 8-cylinder ones.

13. What experimental fact discussed in Chapter 21 makes it safe to predict that gas or Diesel engines that are 80 per cent or 90 per cent efficient will never be built?

14. Compare and contrast the operation of gas and Diesel engines.

15. Why is an ignition system unnecessary in a Diesel engine?

16. Mention the various ways in which the energy in gasoline that is not converted into useful work is used up in the gas engine.

17. Discuss reasons for a Diesel engine's being more economical to operate than a gas engine.

18. The turbo-jet design discussed in this chapter is very general; look up a specific type and be able to describe its operation.

19. Why would rocket propulsion be more satisfactory than turbo-jet propulsion for a trip to the moon?

Multiple Choice

1. In a slide-valve steam engine, reciprocating motion is transformed into rotational motion by means of (a) a slide-valve and an eccentric, (b) a throw and a cam, (c) a piston and a cylinder, (d) a rotor and a stator, (e) a crankshaft and a connecting rod.

2. List all of the following cases in which the principal transformation of energy is in the same direction as that accomplished in a heat engine. (a) Swinging pendulum slowing down because of friction. (b) Nail becoming heated while being driven with a hammer. (c) Boy eating food to provide energy for riding a bicycle. (d) Man rubbing his hands together to warm them. (e) Rocks rising into the air because of a dynamite explosion.

3. In a gas engine that is operating properly, both valves are open during all or almost all of (a) the compression stroke, (b) the exhaust stroke, (c) the power stroke, (d) the intake stroke, (e) no stroke.

4. For any four-stroke-cycle gas engine it is necessarily true that (a) it has four cylinders, (b) it has a compression ratio of 4 to 1, (c) the crankshaft makes four revolutions per cycle, (d) any one piston makes four one-way trips along the cylinder from one exhaust stroke to the next, (e) each valve opens four times and closes four times between successive power strokes.

5. In the turbo-jet engine described in this chapter, the compressor is driven by (a) an auxiliary electric motor, (b) the rotation of the airplane's propeller, (c) an auxiliary compressed-air supply, (d) a separate air motor, (e) the combustion gases as they move toward the jet after burning.

True-False

1. A heat engine may correctly be said to be a device that enables man to violate the Law of Conservation of Energy.

2. The need for more rapid means of transportation furnished the chief impetus for the development of early steam engines.

3. Other things being equal, the higher the temperature of the steam as it enters a steam engine, the more efficient will be the engine.

4. The function of the condensing chamber which Watt added to the Newcomen engine was simply to provide a larger volume of vacuum.

5. A boy riding a bicycle makes use of the crankshaft and connecting rod principle to convert reciprocating into rotational motion.

6. The function of the distributor in a gas engine is to distribute the fuel evenly to all the cylinders.

7. The operation of a crankshaft and connecting rod is better illustrated by a man turning a grindstone than by one operating a Flit gun.

8. In a 16-cylinder automobile engine there are 16 power strokes during every two revolutions of the crankshaft.

9. The German V-2 bomb was propelled by a continuous duct engine.

10. Lack of an atmosphere against which the jet could push rules out any possibility of a rocket ship ever reaching the moon.

Chapter 23

MAGNETISM

MAGNETISM is related to energy because its various manifestations involve forces, and whenever forces act, there exists the capacity for doing work. The facts concerning magnetism are of importance to us in this discussion both because of their direct application in our modern world, and because of the close connection there is between magnetism and electricity.

HISTORICAL BACKGROUND [1]

Knowledge before 1600

No one can say just when, where, or by whom the phenomenon of magnetism was first discovered. The event occurred in whatever part of the world at whatever time some individual first chanced to pick up a piece of naturally magnetic iron ore and noticed that it attracted bits of iron or other pieces of ore. This particular kind of iron ore has the formula Fe_3O_4 and is called *magnetite* or lodestone. That magnetism was known at a fairly early time is evidenced by some of the legends that have grown up about it. One of these tells how it was discovered by a shepherd of Crete, named Magnes. He is said to have experienced great difficulty one day in lifting either his feet or his shepherd's staff from the ground. A little excavation is alleged to have brought to light a large magnetic rock which was strongly attracting the iron nails in his shoes and the iron tip of the staff. According to this legend the word " magnetism " derives from his name. A much more likely theory of the origin of the word is that it came from Magnesia, a city in Asia Minor near which many natural magnets were found.

[1] Much of the historical material on both magnetism and electricity was condensed from *Physics, the Pioneer Science*, Taylor, Lloyd W. Boston: Houghton Mifflin, 1941.

One of the Arabian Nights stories concerns a coastal mountain that was so strongly magnetized that it was not at all unusual for all the nails to be drawn out of the hull of a passing ship, sending it to the bottom. Although none of the stories of this sort can possibly be true because of the fact that magnetic forces are comparatively small, they do indicate the existence of some knowledge of the phenomenon.

Among the earliest verified references is that made by Socrates to iron rings suspended one below the other, the top one being held by a magnet. There is also some evidence to support the claim that as early as the first or second century A.D. the Chinese had learned to suspend a lodestone and use it as a compass for directing them on land journeys. The first use by the Chinese of compasses to guide them on the sea seems to have occurred around 1100 A.D. By that time they had learned how to magnetize long thin pieces of iron by rubbing them with natural magnets. Incidentally, most nations, including ourselves, speak of the compass needle as pointing north, while the Chinese emphasize the other end of it and say it points south. By the latter part of the 12th century the compass was in common use in ships sailing from European countries. The importance that was attached to it is made evident by the penalties which were enforced against anyone who damaged it. One of the codes of the sea provided that anyone caught tampering with the ship's compass should have his hand fastened to the mast by a dagger driven through it, and that the culprit could be released only by tearing himself loose.

Many weird beliefs and superstitions cluttered up the general field of magnetic knowledge in early times. It was believed that contact with garlic would remove the magnetic properties of a compass and elaborate precautions were taken on shipboard to see that no such tragedy occurred. If it did occur, a thorough washing in goat's blood was said to restore the magnet to its original effectiveness. Power to cure all sorts of diseases was attributed to magnetism, as well as the ability to reconcile estranged husbands and wives. Although the second of these two beliefs probably has no supporters today, the first one still has a following. Even in this presumably enlightened age, one can buy coils of wire inside of which the ailing individual is supposed to stand or sit after having plugged the device into a light socket. Many objective investigations of the curative effects of magnetism have been

made during the past 350 years and all have produced exactly the same result, namely, that the only people who derive benefit from such devices are those who sell them.

Magnetic declination and dip, phases of magnetism which we will consider in some detail in connection with the experimental facts of magnetism as a group, were known before 1600. Columbus' writings indicate his knowledge of the former during his voyages to the New World.

The Work of William Gilbert

The modern study of magnetism may be said to have begun with the work of William Gilbert (1540–1603), court physician to Queen Elizabeth and James I of England. In the course of it he collected the magnetic knowledge of his time, including the results of both his own work and those of earlier investigators. Gilbert had two very specific reasons for embarking on a study of magnetism. First, as a physician, he was interested in either verifying or disproving the claims that were being made of the curative effects of magnetism. As we have already pointed out, neither Gilbert nor anyone since his day has been able to find good evidence of the existence of such effects. His second motive grew out of his support of the Copernican model of the solar system. He hoped to find in magnetism a way of accounting for the forces which were necessary to keep the planets in their orbits, as pictured in this theory.

Gilbert's greatest contribution to the knowledge about magnetism was his conclusion that the earth is a huge magnet. This resulted from experiments he carried out with small iron spheres. He magnetized these spheres and investigated the behavior of small compass needles placed at different locations on them. To his surprise he found that the small compasses behaved with respect to his magnetized spheres in exactly the same manner that ordinary compasses behave with respect to the earth. For this reason he was led to conclude that the earth must also be a magnet. Details of the earth's magnetic features will be considered a little later.

In spite of his scientifically sound experimental procedure, Gilbert was not entirely free from the tendency of his day to embellish prosaic observable facts with a certain amount of mystical embroidery.

He tried to explain both magnetic and gravitational forces by attributing to the magnet or the earth a soul which flowed out in the form of invisible radiations that surrounded and drew to itself objects which it attracted. However, Gilbert's speculative imaginings about the " why " of magnetism in no way detract from the importance of his careful investigations of the " how " of it.

With this we will leave the historical background of magnetism and turn to the actual facts about it.

Some Experimental Facts about Magnetism

Our procedure here, as in our discussions of other phenomena, will be in accordance with the steps of the scientific method. We will discuss first a group of typical experimental facts concerning magnetism and then examine a particular theory of the nature of the phenomenon to see how well it fits these facts.

Directional Effect

It is the directional effect, of course, which finds practical application in the compass. Any magnet, whether natural or man-made, if suspended from a string or floated on a cork, will tend to line itself up in a general north and south direction. Under these conditions, the end that points approximately north is called the *north-seeking pole* and the opposite end the *south-seeking pole*. These names are commonly abbreviated simply to " north pole " and " south pole." But the reader may avoid some confusion if he thinks of them by their full names and keeps in mind just how those names were determined.

It will be noted that we spoke of the north-seeking pole as pointing *approximately* north. There are two reasons why it does not point directly north for most places on the earth's surface. One of these is the fact that deposits of iron ore or similar local effects may cause some deviation of the compass needle. The other is more fundamental and concerns the observed fact that the magnetic North Pole of the earth is not at the same location as the geographic North Pole. In fact, the former is some 20 degrees or approximately 1,400 miles from the geographic pole, roughly on a north and south line that would run through Minneapolis, Minnesota. Since the compass needle points at the *magnetic* pole, it obviously will point true geographic north for

only a few places on the earth's surface. The angle between the true north direction and the direction that the compass points for any locality is called its magnetic *declination*. The manner in which this quantity varies over the United States will be evident from Fig. 126. On this magnetic map the lines that run generally north and south pass through places having the same declination. These are called *isogonic*

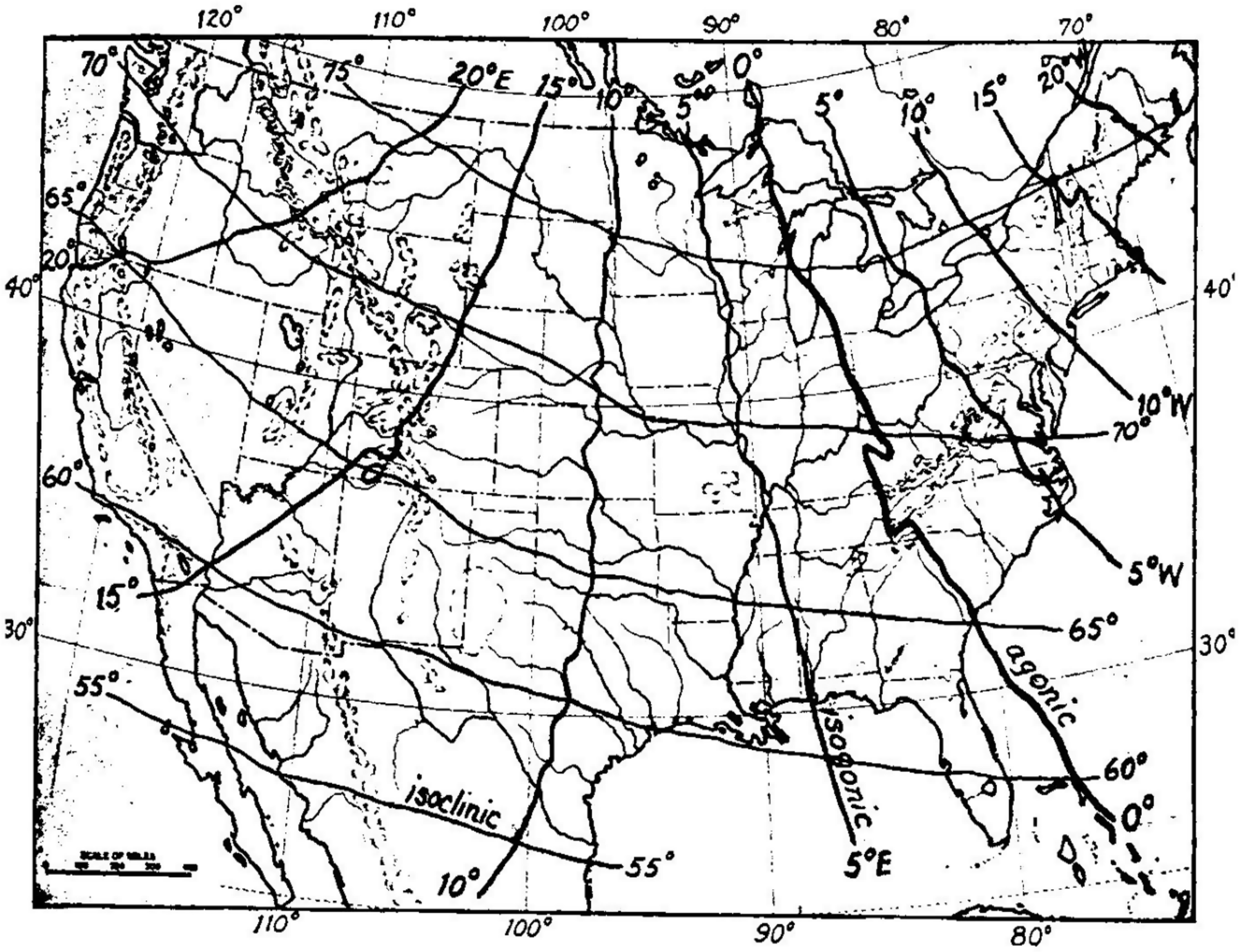


FIG. 126. A magnetic map of the United States for the year 1930, showing the declination of a compass needle from true north and the angle of dip of a dipping needle. The map was drawn from data collected by the U. S. Coast & Geodetic Survey. (From "Classical & Modern Physics," by Harvey E. White, D. Van Nostrand Co., 1940.)

lines. Note that the line of zero declination, that is the line through places in which the compass points true north and south, runs across the middle of Lakes Superior and Michigan, down through Cincinnati and approximately along the boundary between Georgia and South Carolina. All points west of this line have an eastern declination and all points on the east side, a western declination, the extremes in the United States being Maine with 23° W. and Washington state with 24° E.

There is one phase of the directional effect that we cannot observe

with an ordinary compass. In such an instrument, the magnetic needle can only rotate about a vertical axis. Fig. 127 shows a magnetic needle suspended so that it is free to rotate about a horizontal axis. Such a device is called a *dipping needle.* Since the north-seeking pole tends to point directly at the north magnetic pole of the earth, a needle of this sort will not maintain a horizontal position in the northern hemisphere, but will point down, or dip. The angle between the horizontal and the direction such a needle points for any locality is called its magnetic dip. This quantity also varies at different localities as shown by the lines which run across the magnetic map in Fig. 126. These are called *isoclinic lines* and pass through localities having the same dip. Directly over the north magnetic pole, the north-seeking end of the dipping needle would point vertically downward while over the south magnetic pole it would point straight up.

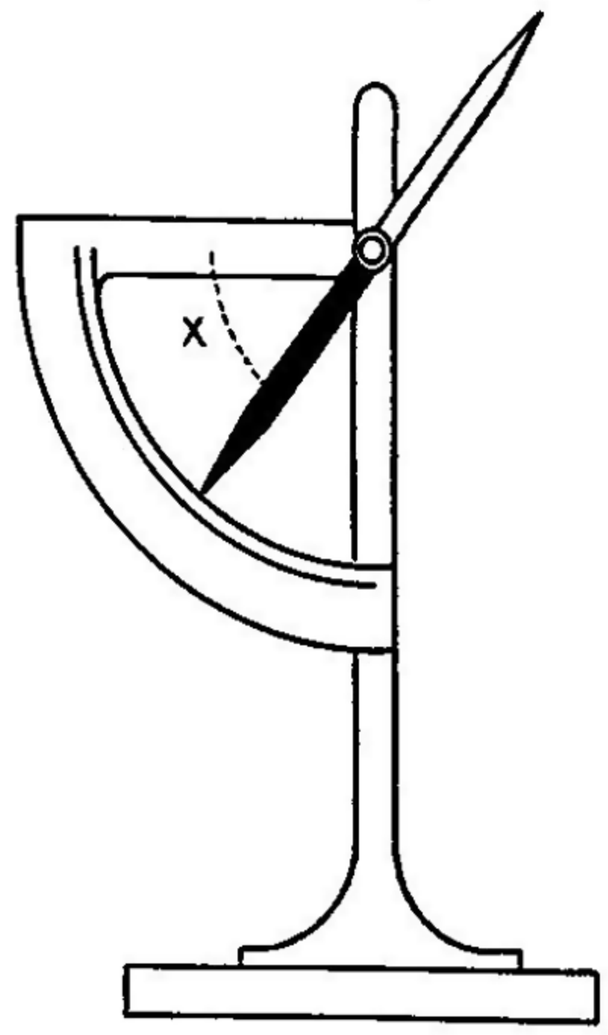


FIG. 127. A dipping needle. Angle X marks the angle of dip.

Attraction and Repulsion

The qualitative facts of attraction and repulsion are pictured in Fig. 128. The first part of the demonstration would be to suspend the magnets freely and determine the north-seeking and south-seeking poles of each. If we assume that to have been done before the compasses and magnets were located as shown in the illustration, then we see that whenever poles of the same kind are near each other there is a force of repulsion between them, while poles of different kinds attract each other. These facts are summed up in the magnetic law which states,

Like magnetic poles repel and unlike attract.

With this law in mind, let us return for a moment to the question of the earth's magnetism. Since unlike poles attract and the north-seeking pole of a compass points north, then we must conclude that the north magnetic pole of the earth, that is, the one up around Hudson

Bay, is a pole of the south-seeking type. Similarly, the earth's magnetic pole that is located down in Antarctica is a north-seeking type of pole. This may appear a bit confusing but is not difficult to keep clear if one remembers the basis on which poles of a compass are named and the magnetic law stated above.

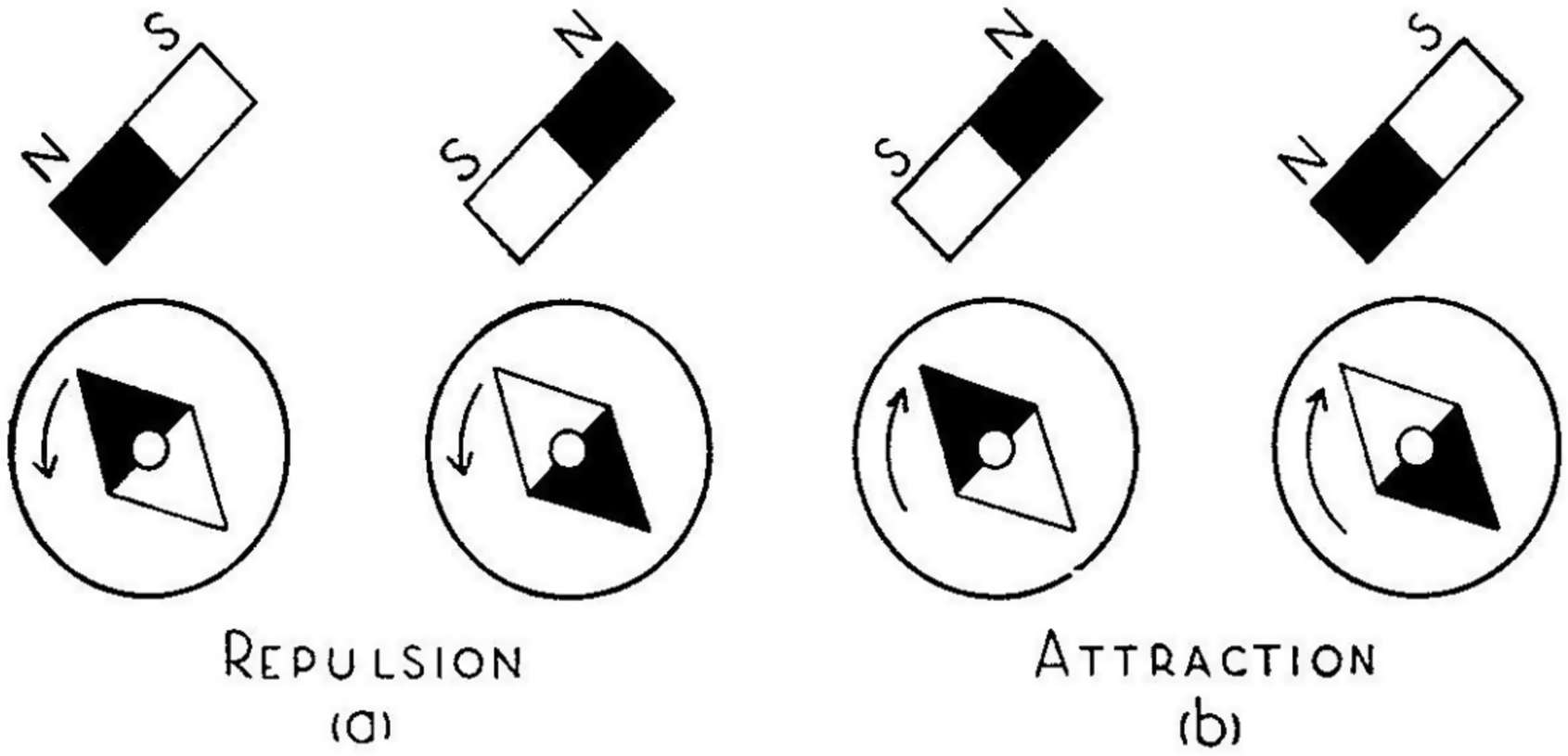


FIG. 128. Diagram indicating forces of repulsion and attraction between like and unlike magnetic poles.

Quantitatively, the force of attraction or repulsion between two magnetic poles is expressed by a relation which bears a close resemblance to the one given in Chapter 20 for gravitational forces. We will state this one also in both words and symbols:

$$\begin{matrix}\textit{Force of attraction} \\ \textit{or repulsion}\end{matrix} = K \frac{\textit{Strength 1st pole} \times \textit{Strength 2nd pole}}{\textit{Distance apart, squared}}$$

or

$$F = K \frac{m_1 \times m_2}{D^2}$$

It is obvious that before such a relation can be used in the calculation of forces one must adopt some sort of basis for measuring the strength of magnetic poles. The one which is used defines a unit magnetic pole as one which repels an identical pole with a force of 1 dyne when the two are 1 centimeter apart in air. When such units are used to measure the strength of the poles and the D is measured in centimeters, then K becomes simply 1 and the force, F, comes out in dynes.

Magnetic Poles in Pairs

We find it to be an experimental fact that it is impossible to have either a north-seeking or a south-seeking magnetic pole all by itself. If a bar magnet is broken into two pieces and each of these broken again into two pieces, the results are as shown in Fig. 129. New poles always appear at the breaks so that each piece is a complete magnet. This holds true no matter how small a section is sliced from one end of a magnet.

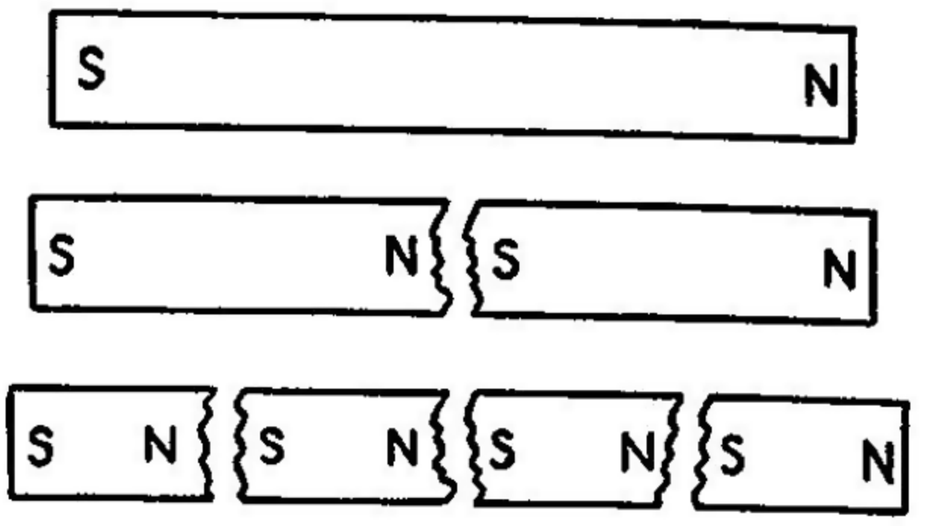


FIG. 129. Diagrams illustrating the effect on the magnetic poles of breaking a magnet into two and then four pieces.

Magnetizing Methods

We will list three methods that may be used to make a magnet out of an unmagnetized iron bar.

One way is to stroke it with a good magnet. This method was employed in making early magnets for ships and in keeping them up to strength. A lodestone that could be used periodically to rub the compass needle was an important part of every vessel's equipment.

A slower way is simply to place a magnet beside the unmagnetized bar and let it stay there for a time. The bar will be found gradually to acquire some magnetism. The process may be speeded up by lightly tapping the bar.

A third method is to heat the bar to a high temperature, red-hot for example, and then permit it to cool while near a good magnet. When it reaches room temperature it will be found to have become magnetized.

Demagnetizing Methods

There are several ways in which a magnetized bar can be made to lose its magnetism. Pounding it vigorously with a hammer will serve the purpose. Another way is simply to heat it red hot and then see that it cools away from all other magnets. Or, if one is in no hurry, the magnet can simply be laid aside for a long time, because magnets gradually lose their magnetism.

The reader will recognize that there must be other methods since when the jeweler demagnetizes a watch he obviously neither pounds it with a hammer nor heats it red-hot. The procedure he uses is to put it in a device that rapidly and repeatedly reverses the magnetism in the watch while constantly reducing it in amount. The watch becomes demagnetized in a manner that may be thought of as somewhat analogous to a swinging pendulum which gradually dies down and eventually comes to rest.

Limit to Extent of Magnetization

Experimentally it is found that there is a limit to the amount of magnetization that an unmagnetized bar will acquire. In other words, as the magnetizing methods described above are applied, there eventually comes a time when the bar that is being magnetized ceases to become any better magnet. At this point it is said to be *saturated*.

Change in Length with Magnetization

It is an observed fact that an iron bar changes slightly in length during the process of being made into a magnet. As the magnetizing procedure begins, the bar first increases a little in length. Soon the change is reversed and it starts to shorten. If it is made into a fairly good magnet, its final length is a little less than it was originally. It should be emphasized that these changes are very small and not at all of the order of magnitude that one might detect with a foot ruler or a yardstick. Nevertheless, they are definitely established and quite measurable by accurate methods. The phenomenon is called *magnetostriction*.

The Molecular Theory of Magnetism

Basic Assumptions

According to this theory of the nature of magnetism, a magnetic material, like iron, is pictured as containing a large number of tiny elementary magnets. When the substance is in a completely unmagnetized state these elementary magnets are thought of as being arranged quite at random. This condition is shown in the upper part of Fig. 130. In this state one would expect no accumulative magnetic forces to make themselves felt at the ends of the bar, since there are

about as many small north-seeking poles pointing one way as the other. In the center picture, the bar is shown in a partially magnetized state. That is, although there are still many elements in a more or less haphazard arrangement, there are more of them pointing toward the left than in any other one direction. The bottom diagram shows the bar completely magnetized or saturated. Here all the elementary magnets are arranged with their tiny north-seeking poles pointing one way and their south-seeking poles the other.

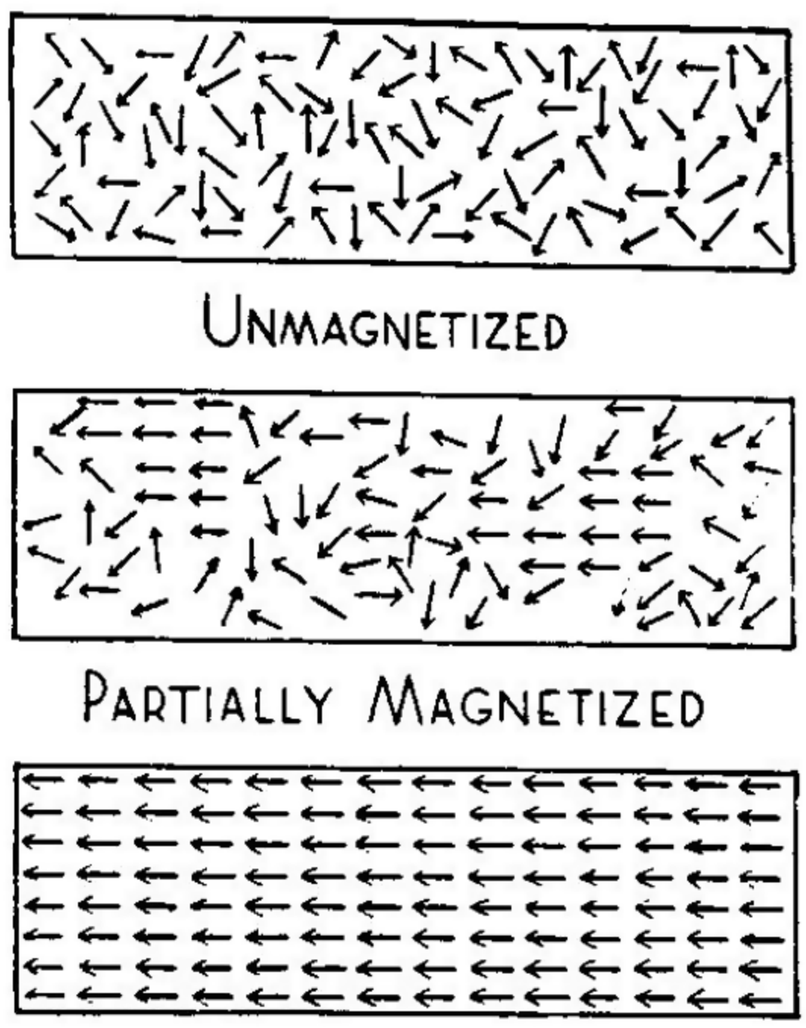


FIG. 130. Schematic diagrams showing arrangement of elementary magnets in an iron bar which is (*a*) unmagnetized; (*b*) partially magnetized; (*c*) completely magnetized or saturated.

Thus, according to the theory, the process of magnetizing a material is not one of adding something to it but rather one of rearranging what is already there. Conversely, the demagnetizing process is one of disarrangement, not one of removal.

To complete the picture of a magnet, as far as we shall be concerned with it in this book, we must say something about the region around it. The space about a magnetized bar seems to possess certain properties not present in that about an unmagnetized one. This may be inferred from the fact that in the first case, bits of iron or compass needles, when in the region, are acted upon by forces, while in the second there are no such effects. The region around a magnet is called a *magnetic field* and the field is thought of as containing *lines of force*, where a line of force is defined as the path along which a free north-seeking pole would travel if one could be turned loose in the region. Since such a pole would be repelled by the north-seeking pole of the large magnet and attracted by the south-seeking one, it will be evident that the lines drawn in the space about the bar magnet in Fig. 131 represent such lines of force. It should be noted that the small compass needles are lined up with these force lines. In fact, it is by the use of small compasses that the field about a magnet may be mapped. The photographs in Fig. 132 show the field arrangement for cases of like

and unlike magnetic poles near each other. The positions of the lines of force in the photographs are marked by iron filings which were scattered about the surfaces on which the magnets are lying. Each filing behaves like a tiny bar magnet or compass needle and orients itself in a direction parallel to a line of force at that particular point.

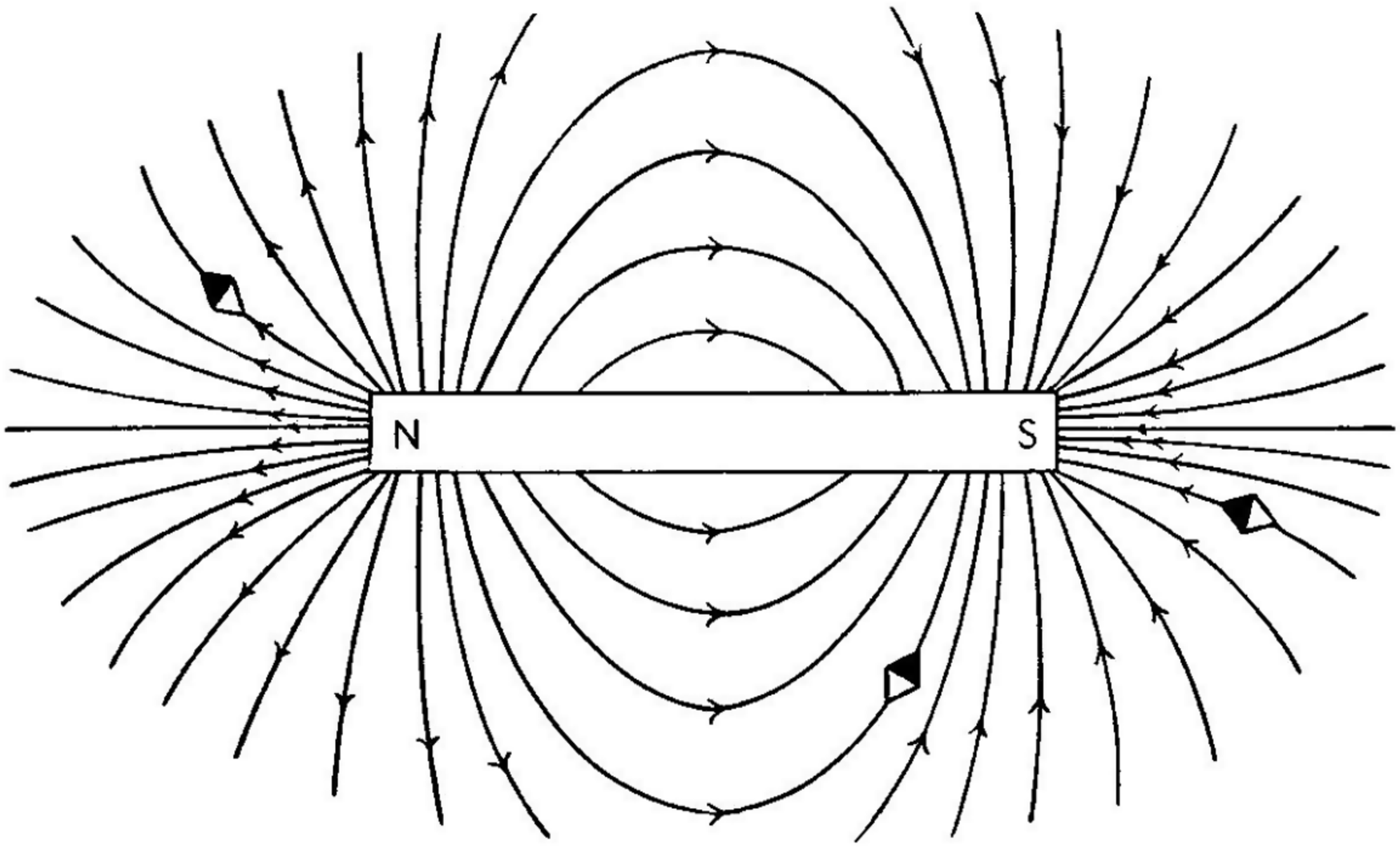


FIG. 131. Diagram showing the magnetic field about a bar magnet.

Above Experimental Facts in the Light of the Theory

Directional Effect: Attraction and Repulsion. — What was called the directional effect earlier in the chapter is seen now to be simply a special case of attraction and repulsion. The north-seeking pole of a compass is attracted by the south-seeking magnetic pole located in the earth up near Hudson Bay. The molecular theory, it is evident, throws no light on why attraction and repulsion occur. It is not, of course, contradicted by any of the experimental facts.

Magnetic Poles in Pairs. — The picture of a magnet as consisting of many small magnets makes reasonable the observed fact that magnetic poles always exist in pairs. If the elementary magnets are of molecular dimensions, then any piece one might cut off the end of a large magnet would contain many of the tiny ones, each of which has two poles.

Magnetizing Methods. — In terms of the theory, any procedure that would tend to line up the elementary magnets in a bar of magnetic material would be expected to impart magnetic properties to the bar as a whole. In view of the facts of attraction and repulsion, it is

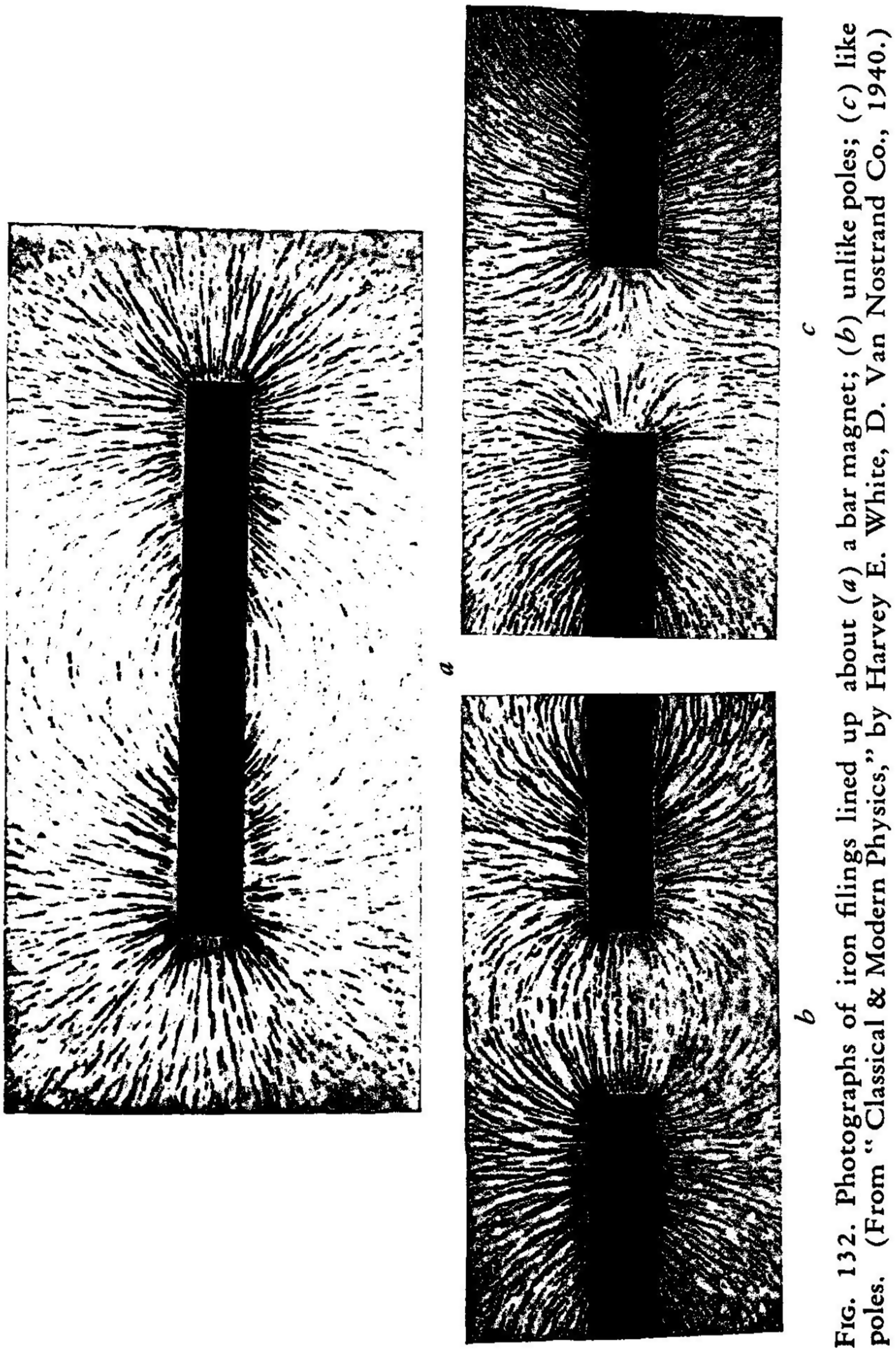


FIG. 132. Photographs of iron filings lined up about (*a*) a bar magnet; (*b*) unlike poles; (*c*) like poles. (From "Classical & Modern Physics," by Harvey E. White, D. Van Nostrand Co., 1940.)

obvious that stroking such a bar with a magnet would tend to do just this. Or placing the bar near a magnet should have some effect, while tapping it should speed up the readjustment process. According to the kinetic theory, heating the bar would put all its molecules in rapid motion. So it might be expected that letting it cool, that is, letting its molecules slow down, near a magnet would result in a final partial magnetic alignment when the bar reached room temperature.

Magnetic Saturation. — Here again, the facts are what one would expect from the theory as stated. If the magnetizing process is simply one of lining up elementary magnets, then a limit would be reached when all the elements were in line. Thus, the phenomenon of saturation is to be expected. Actually, in the manufacture of magnets, saturation is only approximated. Since there is constant molecular motion it is impossible ever to get all elementary magnets in line at the same time and keep them that way.

Magnetostriction. — Recalling the diagrams of unmagnetized and magnetized bars shown in Fig. 130, it will be seen that in the latter case as one goes from one end of the bar to the other, every pole of every elementary magnet is followed by one of the opposite kind. This means that all forces between these small poles from one end of the bar to the other would be forces of attraction. Such is not the case in the unmagnetized bar. Therefore, it is not unreasonable that a bar should be a little shorter when magnetized than when not a magnet. Likewise it does not appear unreasonable that, as the magnetizing process starts, the initial rearrangement of random elements should at first result in a slight lengthening.

Conclusions about the Theory

In the light of the above discussion, the molecular theory of magnetism seems to offer a satisfactory language or picture in terms of which we can describe magnetic phenomena. Although we are limited by the scope of this book principally to qualitative considerations, it may be said that quantitatively also the agreement is reasonably good.

Here again, as in other theories, it should be pointed out that the theory does not " explain " magnetism. To postulate that a big magnet is made up of a large number of little magnets tells us nothing

about why magnetic poles attract or repel each other with forces which can be calculated by the equation given earlier in the chapter. And, as has been mentioned before, to expect it to do so is to misunderstand the function of a scientific theory.

Nothing has been said here concerning the nature of the elementary magnets or their relation to the atoms and molecules of the magnetic material. The reader should keep this in mind while studying Chapter 26. The material considered there may throw light on these points and enable him to suggest a plausible elementary magnet.

Types and Uses of Magnets

So far in this chapter we have referred only to iron as a material from which magnets are made. As a matter of fact there are three so-called " magnetic " elements, iron, cobalt, and nickel. These three are the only ones which exhibit strong magnetic properties. Almost all the rest are affected to some extent by strong magnetic fields but give so slight a reaction that they are commonly called non-magnetic. Formerly, good permanent magnets were made from hard steel, but today most magnets are made from alloys. These usually contain two or all three of the magnetic elements and often small percentages of one or more of the non-magnetic group.

Two properties of magnets are of particular importance in connection with the uses to which they are put. For certain purposes it is necessary to have very strong magnets which will retain their magnetic properties for long periods of time. To cite an example, magnets of this kind find important uses in telephone systems. Such magnets are said to have a high *retentivity*. Hard steel is a material that possesses this property to a considerable degree. The material which probably has a higher retentivity than any other substance known today is the alloy known as Alnico. The composition of this material was discussed in Chapter 17, in connection with our general consideration of alloys.

Substances like soft iron possess magnetic properties essentially the opposite of those just discussed. They can easily be magnetized but lose their magnetism just as readily. These materials are said to have a high *permeability*. The highest properties of permeability known today are attained in an alloy of nickel and iron called

Permalloy. This substance also was discussed in Chapter 17. Magnetic materials with a high permeability are valuable for use in electromagnets. It is necessary that such a device have strong magnetic properties when the electric current is on and just as necessary that there be no attractive forces when the current is off. For instance, an electric crane must be able to pick up the iron at one place and unload it at another. Such a device then should be made of material with a high permeability rather than a high retentivity.

In addition to the uses for magnets that have already been mentioned, they find important applications in telegraph and radio instruments, electrical measuring devices of various sorts, electric motors and dynamos, and many other instruments and machines.

Questions

Discussion

1. Discuss the probable truth of the story of the discovery of magnetism by the shepherd Magnes.

2. Discuss the contributions to the field of magnetism made by Sir William Gilbert.

3. How would you expect an ordinary compass to behave at a place where the magnetic dip is 90 degrees? Explain.

4. Approximately where on the earth's surface would the magnetic dip be zero? Explain.

5. Find what the magnetic declination and dip are for your particular locality.

6. What would you expect the magnetic declination of the geographic south pole to be? Explain.

7. Suppose you were given two iron bars identical in appearance and told that one was magnetized and the other unmagnetized; how could you determine which was which without the use of any other magnets or compasses?

8. How would the force of repulsion between two magnetic poles 20 inches apart compare with that between the same poles when they were 4 inches apart?

9. Describe an experiment which would prove that there are two kinds of magnetism.

10. Draw lines of force about two bar magnets arranged to form a letter T, one magnet being the stem and the other the cross.

11. Name three experimental facts that tend to indicate that magnetizing a bar is not a process of adding something to it.

Multiple Choice

1. The north-seeking pole of a compass placed halfway between the geographic North Pole and the magnetic North Pole will point geographically (a) north, (b) east, (c) west, (d) south.

2. Which one of the following statements describes correctly the manner in which an unmagnetized piece of iron behaves toward a magnet? (a) It will be attracted by both poles. (b) It will be repelled by both poles. (c) There will be no force exerted on it by either pole. (d) It will be attracted by the north-seeking pole and repelled by the south-seeking. (e) It will be attracted by the south-seeking pole and repelled by the north-seeking.

3. List *all* of the following that would tend to magnetize an unmagnetized piece of iron. (a) Pounding it with a hammer. (b) Raising its temperature 40 or 50° F. (c) Heating it red-hot and allowing it to cool in a magnetic field. (d) Swinging it in a circle on the end of a cord. (e) Stroking it with a magnet.

4. The south-seeking pole of a compass located in Washington, D. C. points (a) due south, (b) due east, (c) east of south, (d) due west, (e) west of south.

5. Two magnetic poles 3 feet apart would repel each other with — (a) ⅓, (b) 3 times, (c) 6 times, (d) 9 times, (e) 27 times — as much force as the same two poles if they were 9 feet apart.

True–False

1. The end of a compass that points north in the northern hemisphere will point south in the southern hemisphere.

2. If a bar magnet is broken in two, each piece will be a complete magnet.

3. At the geographic North Pole, one would expect the magnetic declination to be zero.

4. The angle of magnetic dip should be less at the equator than at the southern end of Hudson Bay.

5. Two magnetic poles each having a pole strength of four units would attract each other with a force four times as great as that between poles of two units each, the same distance apart.

6. The name of William Gilbert is associated primarily with the development of the molecular theory of magnetism.

7. Peregrine and Gilbert were contemporaries.

8. Isogonic lines on a magnetic map always run exactly perpendicular to isoclinic lines.

9. The molecular theory of magnetism would lead one to infer that an iron bar should become heavier as it is magnetized.

10. Lines of force in a magnetic field are imaginary lines which represent paths that free north poles would take if released in the field.

Chapter 24

STATIC ELECTRICITY

THE early historical development of man's knowledge of electricity caused him to divide the phenomenon into two categories. He called these *static*, or electricity at rest, and *current*, or electricity in motion. This was somewhat unfortunate because of the fact that later discoveries made it quite clear that the two are identical as far as their fundamental nature is concerned. However, since the terms are still employed in the electrical language and literature of today, we will use them. The reader will avoid confusion if he simply keeps in mind that the difference between static and current electricity lies, not in their nature, but only in the particular electrical phenomena that are included under the terms.

Our procedure in this chapter will be first, to build up an elementary picture of what electricity is, as we think of it today, and then to consider some static phenomena. In the next chapter we will investigate various manifestations of electricity when it flows through wires, that is, when it constitutes an electrical current.

HISTORICAL BACKGROUND

Knowledge before 1600 A.D.

It is interesting to note that essentially all of man's knowledge of electricity has been acquired during the last 300 years. Four phenomena constituted his entire electrical experience some 3,000 years ago and he had no reason to believe that they had any relation to each other. First was the familiar and terrifying lightning, which primitive man interpreted as being either a punishment for past sins or a warning against committing future ones. Second, he was more or less familiar with the fact that certain materials, after having been rubbed vigorously, exerted forces of attraction on light objects. An example

in terms of today's substances is the way in which bits of paper can be picked up by a hard rubber comb or fountain pen after it has been rubbed on one's coat sleeve. Third, people in some parts of the world had made the unpleasant acquaintance of so-called electric fish which, when touched, are able to administer electric shocks of painful proportions. Finally, there was the weird sort of glow that is often visible about the pointed metal tips of ship masts, a phenomenon long known as St. Elmo's fire.

The above four effects were observed by the man of 30 centuries ago and the same four and no more made up the total contact with things electrical for the man of 3 centuries ago.

Development since 1600 A.D.

William Gilbert. — Around 1600 the same William Gilbert who performed the pioneer experiments in the field of magnetism made the first scientific approach to the study of electrical phenomena. His experiments concerned the generation of electrical charges by friction, that is, the second of the four effects mentioned above. He tested a large number of substances to see whether or not, upon being rubbed, they acquired the ability to attract light bodies. Instead of the two or three that had been known previously he discovered the effect could be produced with a great many including amber, diamond, rock crystal, and sapphire. These he called electrics and the phenomenon itself he named *electricity*, both terms coming from the Greek word for amber. Other substances, notably the metals, he grouped as non-electrics. Later experience has shown that Gilbert's non-electrics also can be given a charge by rubbing them, although the charge may be conducted away almost immediately unless the object is insulated from the hand. In detecting and measuring the forces of attraction, Gilbert used a long metallic needle balanced on a point, somewhat like a compass needle, so that it turned easily. But in spite of extensive and painstaking experiments of this sort he never discovered the phenomenon of repulsion either in magnetism or electricity.

Otto von Guericke. — The first machine for the generation of electric charges by friction was devised by the German lawyer, philosopher, and scientist, Otto von Guericke (1602–1686). His machine consisted of a large sulfur sphere mounted on an axle so that it



FIG. 133. The Van de Graaff electrostatic generator. (Photograph courtesy Prof. R. J. Van de Graaff.)

could be rotated with a crank, while the operator's hand rubbed against it. Working with this machine, von Guericke discovered among other things that, under certain conditions, charged objects repelled rather than attracted each other. Many other machines for the generation of electric charges have been built since the time of von Guericke, the most recent being ones designed for the creation of the large voltages

required in certain so-called " atom smashing " experiments. One of the earliest of these, the Van de Graaff generator, is pictured in Fig. 133. The hollow columns are some 25 feet high and 6 feet in diameter. On the top of each column is a sphere of polished aluminum, 15 feet in diameter. Inside each column is a silk belt running from a pulley at the bottom to one within the sphere. Charges are built up on the belts both by friction and by induction, the latter process being one that will be explained later in this chapter. Voltages of the order of 10 million volts can be generated by this device.

Stephen Gray. — The discovery that electric charges can be conducted from one place to another through certain substances was made by an English experimenter, Stephen Gray (1696–1736). Using damp coarse twine as a conductor, Gray was able to conduct an electric charge generated by rubbing a glass tube, over a distance of more than 700 feet. In searching for materials that would serve as nonconducting supports for the twine, he discovered a number of excellent electrical insulators.

Charles Dufay. — The important discovery that there are two kinds of electricity was made by the French scientist, Charles Dufay (1698–1739). He found that the properties acquired by rubbed resinous materials like amber differed in certain respects from those which appeared on rubbed vitreous substances, such as glass. In view of the types of materials involved, the charges on the former he called resinous electricity and on the latter, vitreous electricity. Since the time of Benjamin Franklin the names of resinous and vitreous have been replaced by *negative* and *positive* respectively and it is these that we will use in the rest of our discussion. To indicate the difference between the two kinds, we will describe an experiment similar to the ones carried out by Dufay, although using different materials.

In this experiment, as pictured in Fig. 134, the hard rubber rod in each case has been rubbed with fur or wool and the glass rod with silk, giving them their respective electric charges. Note in (*a*) that the two rubber rods repel each other, while in (*b*) a force of attraction exists between a rubber rod and a glass one, indicating conclusively that the charge carried on one rod is different from that on the other. Finally, in (*c*), two glass rods are seen to repel each other. In other words, there appears to be a relationship between the two kinds

of electricity, similar to that between the two kinds of magnetic poles in that like charges repel each other but unlike ones attract.

In the light of his experimental results, which were of the sort just described, Dufay advanced the first theory of the nature of electricity, a so-called " two-fluid " theory. Its principal postulates were the following: that there were two kinds of electricity, called resinous and vitreous; that they exerted forces of repulsion and attraction in the manner indicated above; that both kinds could travel readily

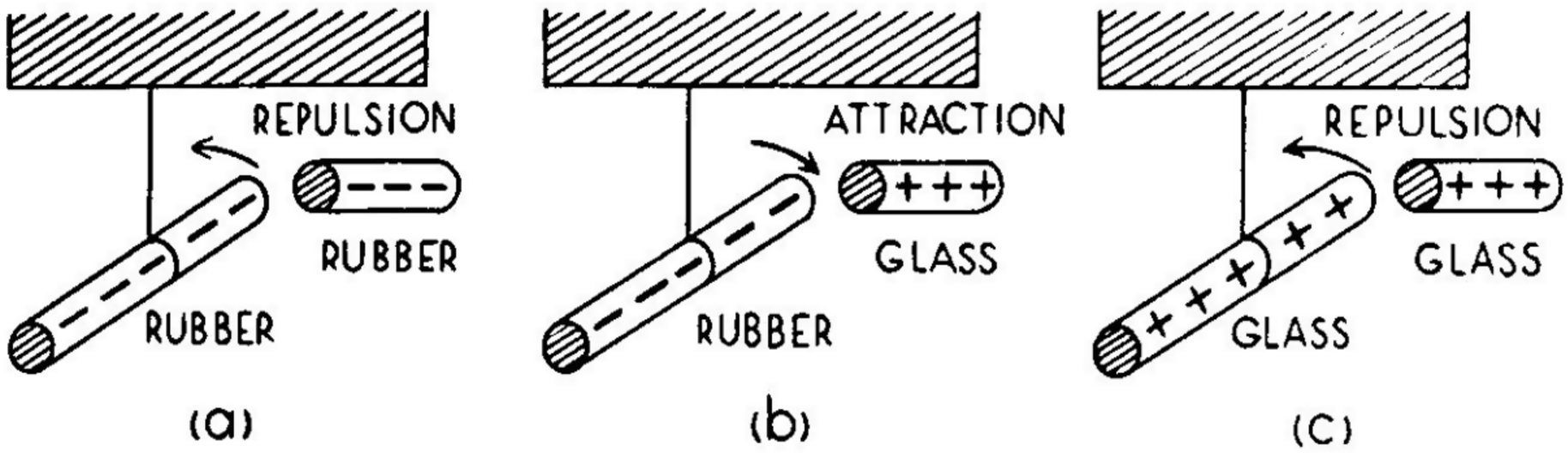


FIG. 134. Diagram illustrating forces between like and unlike electric charges.

through conducting materials; that an uncharged body possessed equal amounts of the two kinds; and that an excess of either kind on an object would give it a net charge of that particular sort. As we shall see a little later, with one exception, these assumptions are in agreement with our present picture, as far as they go.

Benjamin Franklin. — Benjamin Franklin (1706–1790) advanced a one-fluid theory of electricity. He postulated the existence of only one kind, that being the one that Dufay had called vitreous. Then he assumed that there was a certain normal amount or quota of this electricity that any body must have in order to be neutral or uncharged. If it had more than this quota, it acquired properties like that of the glass rod described above, that is, it became positively charged. If it had less than its normal allotment, it was negatively charged. Although later evidence resulted in the abandonment of this theory, the names negative and positive have been kept, as already indicated. Franklin's most important work in this field was his definite identification of lightning as an electrical phenomenon. Other investigators before Franklin had suspected this to be true and two French scientists had even conducted electricity from the atmosphere to the earth by means of long pointed rods. Franklin was not certain that in these cases the electricity had actually come from the clouds

and devised his kite experiment in order to use the storm cloud itself as the source. The experiment is well known to everyone. Its result was that Franklin was able to show that what he collected in a Leyden jar, or condenser, using the storm cloud as a source, had exactly the same properties as electricity collected in a similar jar from an ordinary electrical machine. One of the most remarkable features of the experiment was the fact that Franklin sustained no injuries while performing it. Several European scientists who tried to repeat it were severely shocked and one was killed. Franklin's work marked about the last that was done with what might be called purely static phenomena. The development since his time has largely involved electric currents or combinations of the two sorts of manifestations.

Present Picture of the Nature of Electricity

The facts already mentioned play an important part in the formation of the modern theory of what electricity is. In this section we will try to summarize and bring together these and more recently established facts, all of which enable us to make up a kind of model of this thing we call electricity.

Kinds of Electricity

Today we talk in terms of two kinds of electricity, to which we give the names negative and positive. One way of defining them is with regard to the materials on which they may be generated. We may define negative electricity as the kind that appears on rubber, amber, or similar substance when it is rubbed with wool or fur, and positive electricity as the kind that appears on glass or a like substance when it is rubbed with silk. It should be pointed out that rubber will acquire a negative charge and glass a positive charge regardless of what kind of cloth is used as the rubbing material. The materials mentioned simply seem to give the best results.

The Nature of Negative Electricity

One of the most important discoveries of the last half century was the fact that electricity is what we may call " atomic " in nature. This means that there is a fundamental quantity or unit charge of electricity which cannot be divided into smaller amounts. Thus any

electric charge, regardless of its size, consists of an exact, whole number of these units. Some of the evidence for this fact, which applies to both positive and negative electricity, will be considered in Chapter 26. The unit of negative electricity is the *electron*. In view of the fact that the electron has mass, it was formerly thought of as being a particle of matter which carried an electric charge. But since modern evidence quite definitely implies that in the last analysis matter is electricity, today the electron is commonly described as itself being the basic unit of negative electricity. So the electron's charge may be said to be —1. Its mass has been found to be very small, only about 1/1850 that of a hydrogen atom. Hydrogen, it will be recalled, is the lightest of all the elements. In more conventional units the mass of the elec-

tron equals .000,000,000,000,000,000,000,000,000,9 grams, or
9×10^{-28} grams.

The Nature of Positive Electricity

Up until 1932 the *proton* was the only candidate for the honor of being the fundamental unit of positive electricity. Its charge is +1, that is, it is equal in size to that of the electron but opposite in kind. Its mass, however, is many times that of the electron, being about equal to that of the hydrogen atom. In fact, as we shall see a little later, the hydrogen atom may be thought of as consisting of a particular kind of arrangement of one proton and one electron. By about 1920 physicists were hopeful of being able to prove that atoms of all the elements were composed simply of varying numbers and groupings of electrons and protons. In recent years several other fundamental building blocks have been discovered which somewhat complicate this simple picture. The one that is related to our consideration of the nature of positive electricity is the *positron*, discovered in 1932. Its charge is identical with that of the proton while it has the same mass as does the electron. Therefore, the proton can no longer be said to be the only unit of positive electricity. However, in view of the fact that the positron only comes into being under highly specialized conditions and then does not remain free for any appreciable length of time, we will employ only electrons and protons for our discussion in the balance of this chapter.

The roles played by these units in the building up of atoms will be considered in Chapter 26.

Mobility of Electrons and Protons

It is in this connection that the modern picture differs from that of Dufay. Whereas he thought of both kinds of electricity as traveling readily through conducting materials, the evidence today indicates that only electrons exhibit this freedom of movement. Protons, on the other hand, remain relatively stationary, vibrating about fixed positions but not moving bodily through conductors. To illustrate this someone has suggested an analogy of a field containing a rabbit and a tethered cow. With respect to relative freedom of motion, the former is analogous to the electron and the latter to the proton.

Attraction and Repulsion

We have already pointed out that the relationship which expresses the qualitative facts in this connection is similar to the one for magnetism, and so, may be stated,

Like electric charges repel; unlike charges attract.

Quantitatively also there is a similarity between the two cases. The magnitude of the force between electric charges is given by,

$$\begin{matrix}\textit{Force of attraction}\\ \textit{or repulsion}\end{matrix} = \mathrm{K}\ \frac{\textit{1st Charge} \times \textit{2nd Charge}}{\textit{Distance apart, squared}}$$

or

$F = \mathrm{K}\,\dfrac{q_1 \times q_2}{D^2}$ where q_1 and q_2 are the two electric charges and D is their distance apart. K here has the same significance it had in the magnetic equation.

The close resemblance between the form of the expressions for gravitational, magnetic, and electrical forces has sometimes led people to draw the erroneous conclusion that the phenomena may also be essentially identical. As a matter of fact, the resemblance is simply one of form. For example, there is no such thing as gravitational repulsion which sets it apart from the others. Then between electricity and magnetism, although the two are related, there are important differences. For instance, the process of magnetization is one of arrange-

ment of something already present while that of charging is one of adding or removing something. Also in the electrical case the charge can be removed by connecting a wire from the object to the earth while demagnetization cannot be accomplished in this fashion.

Co-existence of Opposite Charges

In proving the existence of two kinds of electricity earlier in the chapter, we confined our attention to the charges on the rubber and glass rods. Now let us ask also about the fur, wool, and silk. Experimentally we find that when the rubber rod is given a negative charge by rubbing it with fur, at the same time the fur acquires a positive charge *of the same size*. Similarly, in the glass and silk case, the silk becomes negatively charged to exactly the same extent that the glass becomes positively charged. This means that it is impossible to generate an electric charge of one kind without at the same time generating one of the opposite sort. Here again, it will be noticed, is a similarity to magnetism. In that field we found that magnetic poles always occurred in pairs.

Some Static Phenomena in the Light of the Above Picture

Now, having built up a picture of the nature of electric charges and their properties, we will ask just how certain static phenomena fit into the theory.

Charged and Uncharged Bodies

Since the charge on the proton is equal in size but opposite in kind to that on the electron, it is obvious that equal numbers of the two would cancel each other out. In other words, a neutral or uncharged body, in terms of our picture, is one which possesses equal numbers of protons and electrons. To have a negative charge then, an object must possess an excess of electrons over protons. If it possesses fewer electrons than protons it is positively charged.

In terms of the process of charging, since only the electrons are able to move about freely, to give a neutral body a negative charge electrons must be added to it, while to charge it positively electrons must be removed from it. In the case of hard rubber and fur, both

neutral to begin with, the rubbing process transfers electrons from the fur to the rubber. Rubbing glass with silk, on the other hand, causes a migration of electrons from the rod to the cloth.

Either negatively or positively charged bodies can be discharged, or made neutral, by connecting them to the earth, a process known as "grounding." This can be accomplished by actually connecting the body to the earth with a conducting wire or by simply touching the charged object with one's hand. Grounding a negatively charged body makes it possible for the excess electrons to flow into the earth, which may be thought of as a sort of vast reservoir from which electrons can be taken or into which they can be put more or less at will. If a positively charged body is grounded, it becomes neutral because electrons flow on to it from the earth in sufficient quantities to just make up the electron deficit.

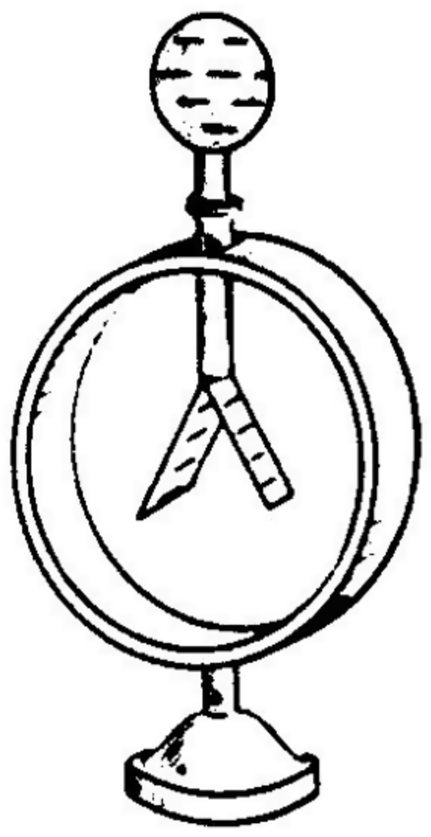


FIG. 135. A negatively charged electroscope.

In this connection it is desirable that we describe the instrument that is commonly used to detect the presence of an electric charge on a body. Such a device is called an *electroscope*. The charged rod suspended in the manner shown back in Fig. 134 is a simple form of electroscope. A more commonly used type is the gold-leaf electroscope such as is pictured in Fig. 135. It consists chiefly of a metal rod, on one end of which is a metal knob while suspended from the other end is a pair of gold foil leaves. The lower part of the rod and the leaves are enclosed in a glass container to protect them from air currents. When it is uncharged the leaves hang vertically downward. But if either kind of charge is given to the instrument the leaves will separate because of the force of repulsion between the like charges on them. When it is used to detect the presence and kind of charge on an object, the electroscope is usually charged one way or the other. This may be done by contact with a charged rubber or glass rod. Suppose we wish to so use it and have charged it negatively as shown in the diagram. If the unknown object is uncharged, its approach will have no effect on the leaves. If it carries a negative charge, it will repel more electrons down into the leaves as it comes near the elctroscope and so cause them to separate still farther. An approaching positively charged

body will attract some electrons up into the knob causing the leaves partially to collapse.

Electrostatic Induction

In discussing the phenomenon of electrostatic induction we will make use of the electroscope described above. Part (*a*) of Fig. 136 shows such a device in an uncharged condition. This means, as previously pointed out, that there are just about the same number of electrons as protons on it. Now if we bring a negatively charged object

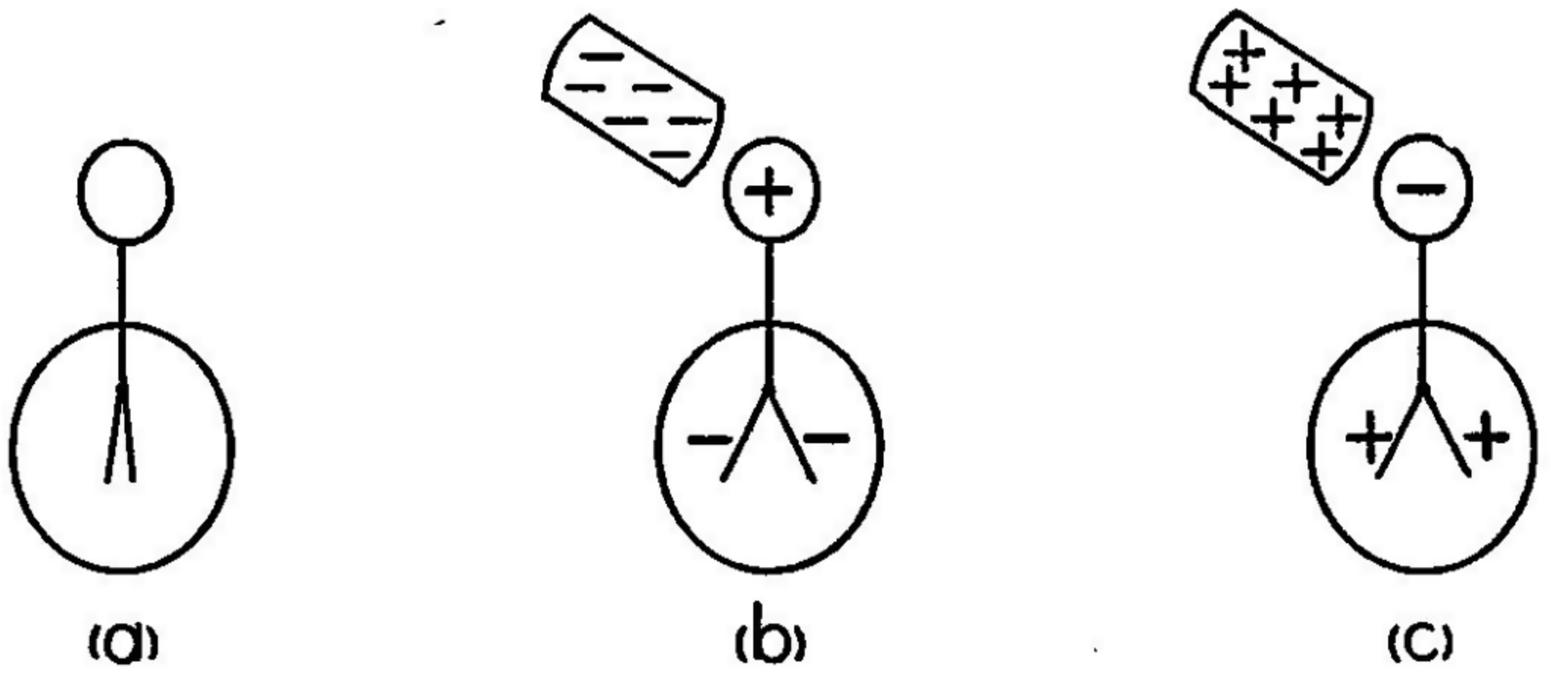


FIG. 136. Diagram illustrating meaning of electrostatic induction.

near the knob, but not touching it, we find that the leaves diverge because of the fact that electrons have been repelled down into them from the knob. Thus, as shown in (*b*), although the instrument as a whole has equal numbers of electrons and protons, their uniform distribution has been interfered with in such a way as to give the knob a positive charge and the leaves a negative charge. The result of bringing a positively charged body near the uncharged electroscope is indicated in (*c*). In this case electrons have been attracted up out of the leaves. Charges produced in this fashion are said to be induced, and the process is called *electrostatic induction*, in contrast to charging by conduction which involves actual contact between the charging body and the electroscope knob.

In the cases just described, when the charged object is removed from the vicinity of the electroscope, the electrons on the latter immediately resume their original distribution and the instrument again is uncharged. The four parts of Fig. 137 show how an electroscope may be given a permanent charge by induction. In (*a*), a negatively charged body has been brought near the knob of the electroscope, repelling electrons down into the leaves. Grounding the instrument

with his hand in (*b*), the demonstrator provides a path of escape to the earth for the excess electrons on the leaves, causing them to become neutral, while the knob remains positively charged. In (*c*) the hand has been removed, eliminating any connection with the earth. When the original charged object has been removed in (*d*), the electrons which are left distribute themselves as uniformly as possible

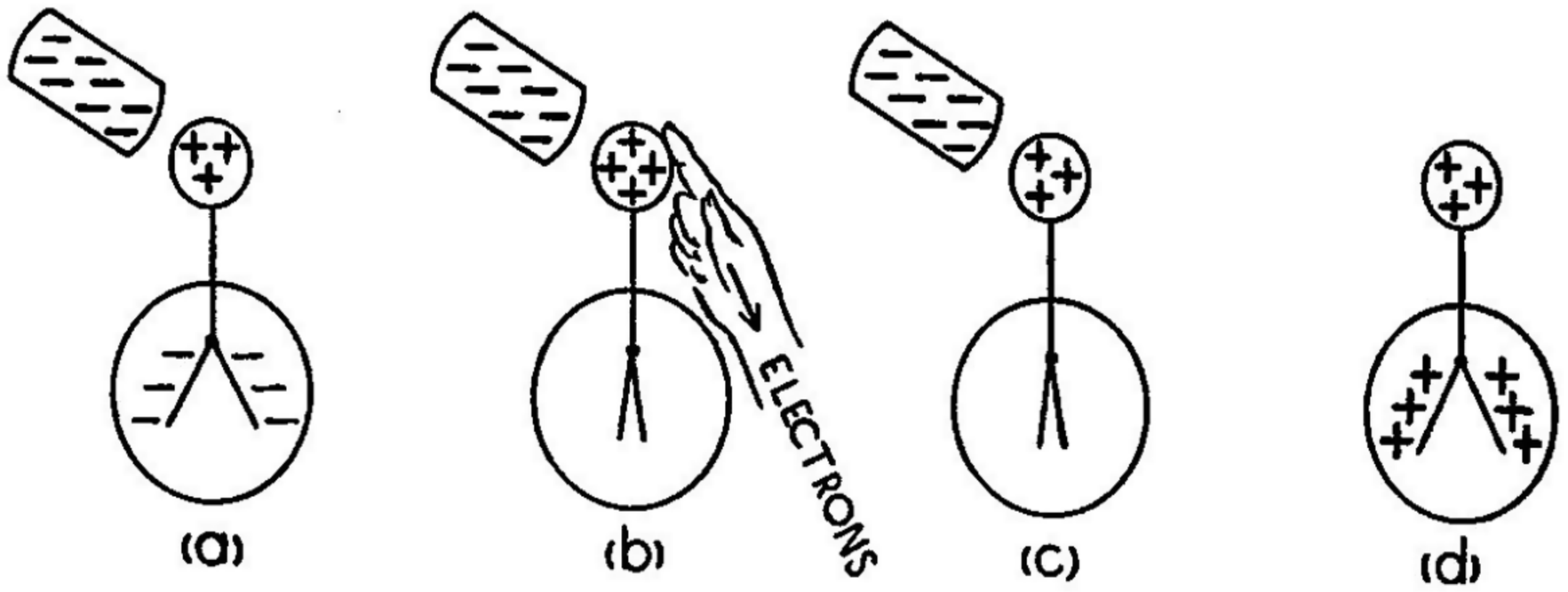


FIG. 137. Diagram illustrating the method of giving an electroscope a permanent positive charge by induction.

throughout knob, stem, and leaves. But since some escaped in (*b*) there are now fewer of them than there are protons, so that the instrument as a whole is positively charged. To check his understanding of this process, the reader should now figure out just how he would go about giving an electroscope a negative charge by induction.

Static Discharges

We have already described how an electrostatic charge can be conducted off to the earth through a wire or through one's body. There are two additional ways in which charged bodies become neutral, one quite peaceful and gentle and the other more or less violent.

It happens that because of the repelling force between like charges, any charge on an object tends to spread out over its surface. If the object is flat or uniformly curved this distribution is uniform, but if it has any sharp points there is a concentration of charge near these which tends to escape or " leak " off. Actually, of course, the leaking off or neutralizing process consists in the escape of electrons if the charge on the body is negative or in the arrival of electrons if it is positive. Such a discharge often produces a visible glow of light when it occurs at the tips of flag poles, ship masts, or lightning rods, and is the phenomenon referred to earlier as St. Elmo's fire.

The value of a lightning rod in preventing a building from being struck involves this sort of discharge. Suppose a positively charged cloud drifts over a building protected by lightning rods. The positive charge on the cloud will induce a negative charge on the building and earth beneath it. But if the lightning rods have sharp points and are well grounded in damp earth, the charge can escape in the manner

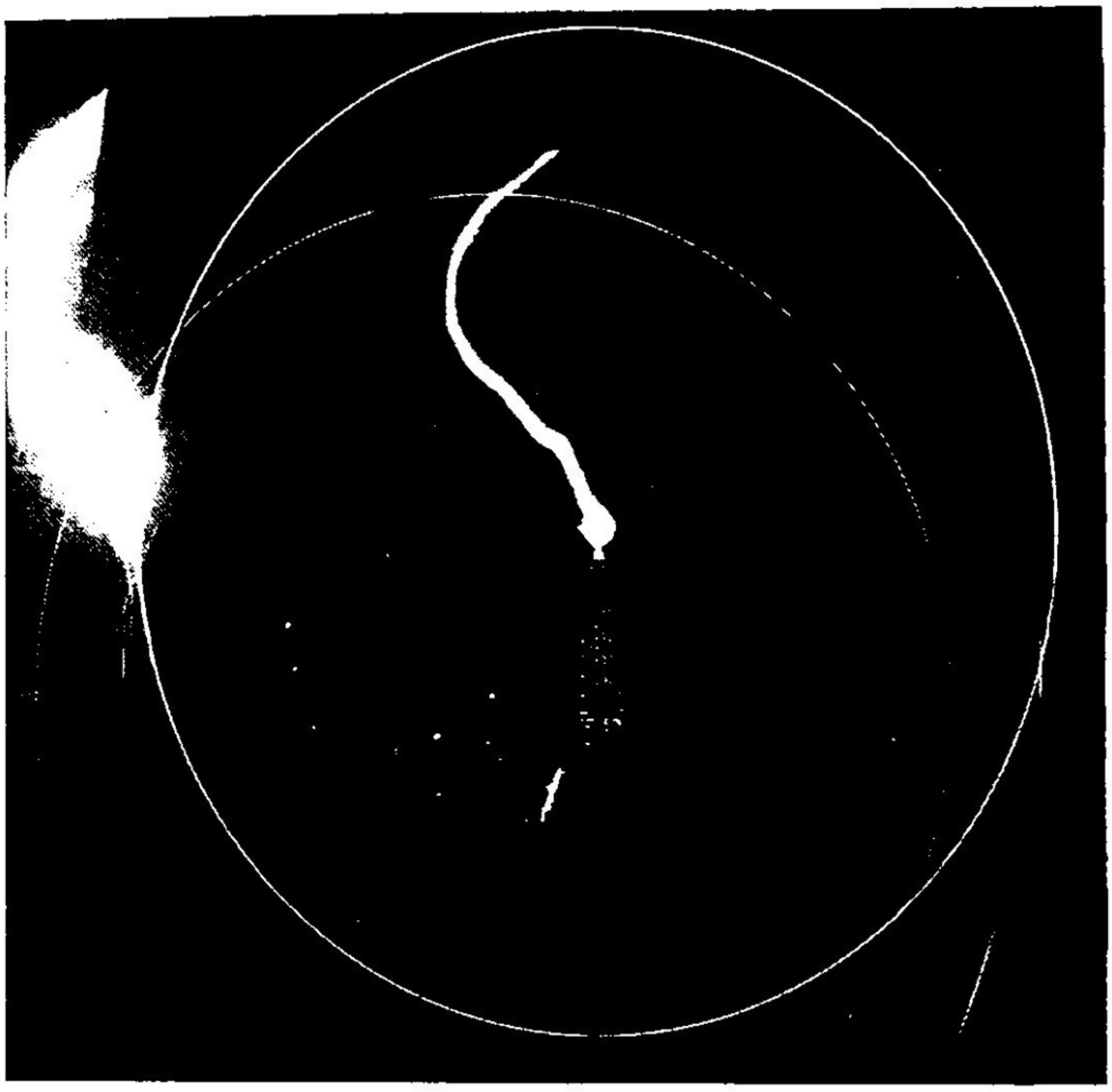


FIG. 138. Lightning striking the Empire State building in New York City Photograph taken August 6, 1940 by Paul C. Edwards. (General Electric Company.)

described above, before a disrupting stroke occurs. In case the flash occurs anyway and the building is struck, good lightning rods still provide protection because they offer a path of low resistance into the earth for the electricity to follow. If they are not well grounded however, lightning rods are an additional hazard rather than a protection.

Lightning flashes are examples of the more violent type of static discharge referred to above. Such a flash, or indeed any electric spark occurs when the opposite charges become so large that the force of attraction is sufficient to cause a sudden and disruptive transfer of elec-

trons through the non-conducting material between the charges. It should be noted that in every case of discharging a charged object, electrons move from one place to another.

Lightning striking the Empire State building is shown in Fig. 138. The steel framework of such buildings is well grounded and serves to conduct the charge directly to the earth. This is the reason why, although modern skyscrapers are frequently struck with lightning, no one has ever been injured by such a stroke while in a building of this sort.

In spite of the very considerable damage that may be done to poorly protected buildings by a stroke of lightning, the actual amount of electricity involved in one flash is rather small. It has been shown that the quantity of electrical energy stored up in an average thunder cloud would be sufficient to operate a medium-sized light bulb less than one minute.[1] However, the total amount of electricity discharged in all thunderstorms taken together is considerable since between 10 and 20 million of these meteorological disturbances occur on the earth every year.

Questions

Discussion

1. Name the four electrical phenomena that man knew about some thousands of years ago.

2. Describe the contributions of each of the following to our knowledge of static electricity:

a. Otto von Guericke
b. Charles Dufay
c. Benjamin Franklin
d. Stephen Gray
e. William Gilbert

3. Find out some additional information about the construction and operation of the Van de Graaf electrostatic generator.

4. Distinguish between Dufay's 2-fluid theory of the nature of electricity and Franklin's 1-fluid theory.

5. Contrast the electron and the proton as to charge and mass.

6. Name four experimental facts about electricity and show to what extent they are in qualitative agreement with the picture of the nature of electricity presented in this chapter.

7. Describe methods of charging a body negatively and positively by both conduction and induction.

[1] Schonland, B. F. J., *Scientific Monthly*, April, 1936, pp. 368–369.

8. Draw a diagram of an electroscope and explain, in terms of the movements of electrons, just how it may be given a permanent charge without the charging body coming into contact with it.

9. Explain just how a lightning rod is a protection to a building and show why it should be well grounded.

10. Give reasons for the statement that a building with a steel frame is a comparatively safe place to be during a severe electrical storm.

Multiple Choice

1. An object contains 11 million electrons and 10 million protons; to have it positively charged one would (a) leave it as it is, (b) add at least a million electrons, (c) add more than a million protons, (d) remove more than a million electrons, (e) connect it to the ground.

2. A neutral or uncharged body may be defined as one which possesses (a) neither electrons nor protons, (b) equal numbers of electrons and protons, (c) more electrons than protons, (d) fewer electrons than protons.

3. List *all* of the following that are in agreement with the picture of electricity as presented. (a) 2,000 electrons would weigh a little more than 1 proton. (b) A proton carries between 1,800 and 1,900 times as large a charge as does an electron. (c) Wool that has been rubbed on hard rubber attracts fur that has also been rubbed on hard rubber. (d) During a lightning flash between a positively charged cloud and negatively charged earth, electrons move from the earth to the cloud. (e) By very vigorous rubbing of silk on glass, the former can be given a larger charge than the latter.

4. A 2-fluid theory of the nature of electricity was first advanced by (a) Benjamin Franklin, (b) William Gilbert, (c) Otto von Guericke, (d) Stephen Gray, (e) Charles Dufay.

5. Included in the electrical phenomena that man had observed 600 years ago was (were) *one or more:* (a) the ability of some materials to attract light objects, (b) repulsion of like charges, (c) relative weights of electrons and protons, (d) the mobility of electrons, (e) the existence of two kinds of electricity.

True–False

1. Because of their iron and steel framework, modern skyscrapers are particularly dangerous places to be during a thunderstorm.

2. If the leaves of a positively charged electroscope move farther apart when an object is brought near it, the object is charged positively.

3. An electroscope that was charged permanently by induction from fur that had been rubbed on amber, would acquire a negative charge.

4. Non-conductors like glass or rubber contain no electrons.

5. William Gilbert never learned that under certain conditions some charged bodies would repel rather than attract each other.

6. Copper was classified by William Gilbert among what he called non-electrics.

7. The name of Van de Graaf is associated with the discovery that charges can be conducted through metal wires.

8. A combination of 1,000 electrons and 1,000 protons would weigh about the same as 2,000 hydrogen atoms.

9. The force of attraction between two unlike charges 3 feet apart would be 4 times as great as that between the same two charges when they were 6 feet apart.

10. Although lightning rods are a protection when a building is struck, they can have no effect toward keeping it from being struck.

Chapter 25

CURRENT ELECTRICITY

THE applications of current electricity to modern life are so numerous and varied that it is obviously impossible to consider them exhaustively in this book. Therefore, we shall confine our discussion to four main points. First, it is desirable that we establish the connection between electric currents and static electricity. Second, we shall consider the essential features of several methods of generating electric currents. Third, we shall look at certain important characteristics of an electric circuit and finally, summarize some of the ways in which electric currents are employed.

Relation to Static Electricity

In considering static electricity, we discussed several ways in which a charged body can be made neutral, and pointed out the fact that every one of the methods involves the transfer of electrons from one place to another. The manner of this transfer varied from the sudden disruptive transfer characteristic of a spark discharge, like lightning, to the less violent movement of electrons which occurs when a charged electroscope is connected to the earth by a conducting wire. In the latter case, we have an electric current which differs from the ordinary current in that it flows for a very short period of time, while that produced by a battery or dynamo can be maintained indefinitely.

In other words, if great numbers of electrons move riotously, instantaneously, and disruptively from a negatively charged body to a positively charged body, it is a spark discharge, but if they travel through a conducting wire in a somewhat more dignified, parade-like fashion, we call it an electric current.

Methods of Generating an Electric Current

Since an electric current consists of a flow of electrons through a wire, the production of such a current requires simply that one create conditions such that forces will be exerted on electrons to make them move from one place to another. Several procedures have been developed which tend to give electrons a migratory urge of this sort.

Generation by Chemical Action

Historical Background. — The first method used in producing an electric current was chemical in nature. Credit for its discovery is given to an Italian physician named Luigi Galvani (1737–1798). One day while engaged in dissecting a frog, Galvani noticed a contraction of the leg muscles whenever a nearby electric machine was in operation. Further investigation showed that the same twitching effect could be obtained by simply connecting the nerve and muscle of the leg to dissimilar metals. But no such result was obtained if only one metal was used or if non-conductors were employed. There were obviously two possible sources of the phenomenon. Either the current was set up at the junction of the two metals or it was a property of the animal tissues. Galvani favored the latter view and in 1791 announced his discovery, attributing the current to what he called " animal electricity " or as it came to be known, " Galvanism." In this connection, Galvani is an excellent example of a scientist who behaved most unscientifically with regard to a hypothesis which he himself had advanced. He became so prejudiced in favor of his animal magnetism theory that he was quite unable to view objectively later evidence which definitely contradicted it and finally caused it to be discarded.

Another Italian, Alessandro Volta (1745–1827), a professor of physics in the University of Pavia, established the true source of the electric current. He demonstrated that it could be produced by the action of dissimilar metals without the presence of animal tissue of any sort. In the course of his experiments he developed the first electric battery, a device known as a *voltaic pile*. Although he tried a number of different materials he found that the best results were obtained when he used silver and zinc as the two metals. The pile consisted of a series of small discs of these metals and of cardboard, the

latter kind having been soaked in a salt solution. Then he piled the discs up one on another, in the order, silver, zinc, cardboard, silver, zinc, cardboard, and so forth, ending with zinc. By connecting wires to the top and bottom discs he was able to get continuous electric currents which were of substantial size.

All the essentials of a modern electric cell or battery were present in the voltaic pile. Developments since that time have been largely directed toward making cells more convenient to use and toward eliminating various undesirable chemical reactions.

The Simple Voltaic Cell. — As an example of the way in which chemical energy can be converted into electrical energy, we will con-

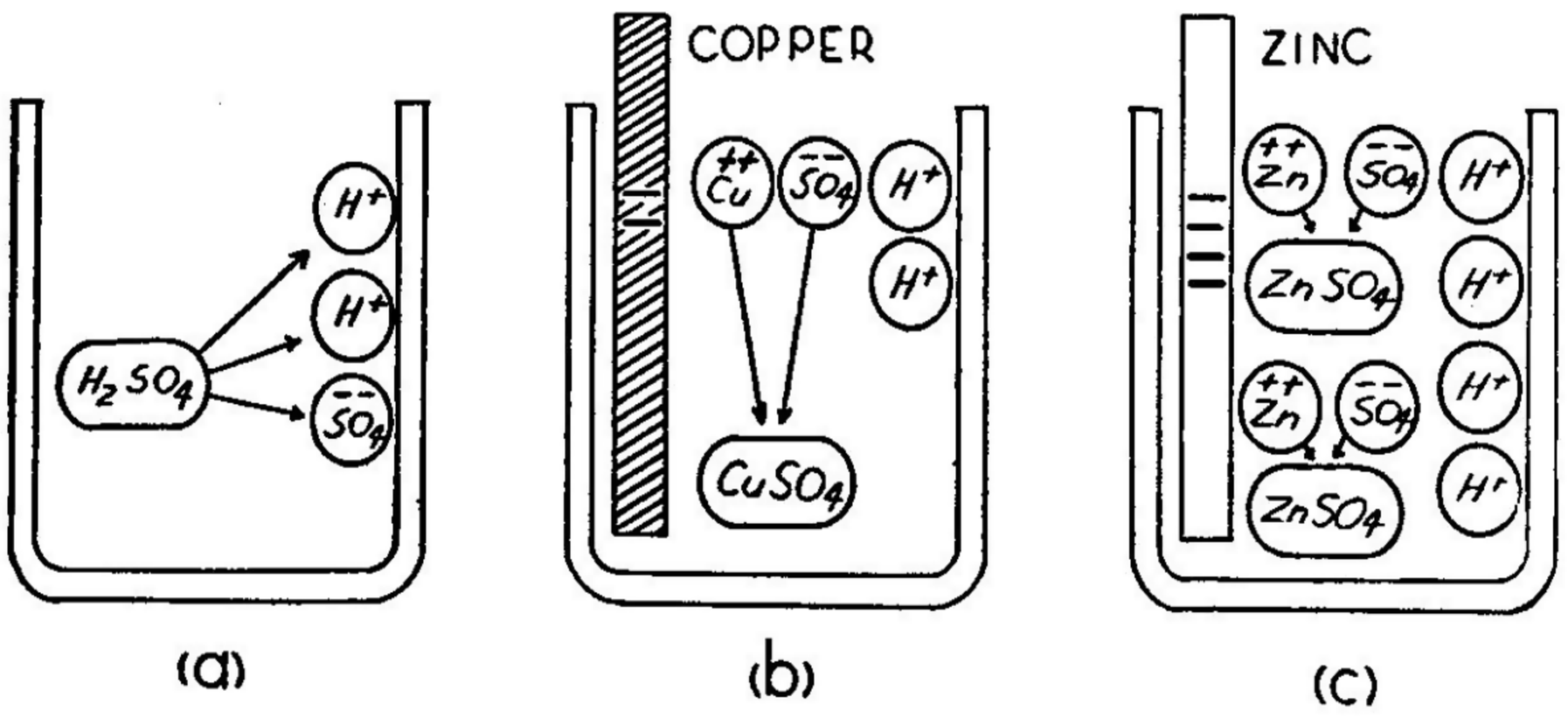


FIG. 139. (*a*) Ionization of sulfuric acid. (*b*) Action of copper in sulfuric acid. (*c*) Action of zinc in sulfuric acid.

sider a particular type of voltaic cell. Before starting to trace the steps in the action of this cell, let us emphasize again that to produce an electric current, one must simply create a set of conditions such that an urge is provided for electrons to move from one place to another through a conducting wire. Then, as they travel along the wire, we are able to use them to ring bells, light lights and so forth.

This voltaic cell consists of plates of copper and zinc immersed in a solution of sulfuric acid in water. Now when sulfuric acid (H_2SO_4) is placed in water the acid molecules break up into what are called H ions and SO_4 ions, each molecule producing two of the former and one of the latter. Each H ion consists of a hydrogen atom which, in breaking away from the acid molecule, left one electron behind. Therefore, in line with our discussion in the preceding chapter, it has a net charge of $+1$. To indicate this we will designate an H ion as

H^+. Since the SO_4 ion represents the portion of an acid molecule that is left after two H^+'s have broken away, it can be called SO_4^{--}. This breaking-up process, which is shown schematically in (*a*) of Fig. 139, is called ionization or electrolytic dissociation.

If a copper plate is dipped into this liquid, as shown in (*b*) of the figure, copper ions begin to go into solution. Each ion consists of a copper atom that has left two electrons back on the plate and so may be designated by Cu^{++}. This tendency to go into solution is called "solution pressure," a quantity whose value varies for different metals. Because of the electrons left behind, the plate becomes increasingly negatively charged as the dissolving process continues and this, since unlike charges attract, tends to hold back the copper ions. In other words, the dissolving process itself results in the setting up of a force which opposes the solution pressure. It will be noted in the diagram that the Cu^{++} ions unite with the SO_4^{--} ions to form copper sulfate ($CuSO_4$).

Part (*c*) of the diagram indicates that when a piece of zinc is placed in the sulfuric acid solution the results are quite similar to those just described for copper. The important difference is that the solution pressure of zinc is greater than that of copper, which means that a larger negative charge is built up on the zinc than on the copper before the opposing forces described above balance.

Now let us turn to the case where plates of both copper and zinc are immersed in a solution of sulfuric acid in water, that is, to the simple voltaic cell. The action, pictured in Fig. 140 can most clearly be described by breaking it down into a series of separate steps. However, it must be remembered that when the cell is in operation, these steps are occurring at the same time, not consecutively.

1. The sulfuric acid solution dissociates forming H^+ ions and SO_4^{--} ions in the proportions mentioned above.
2. The solution pressure of zinc being greater than that of copper, the net effect is that zinc goes into solution as Zn^{++} ions and copper does not dissolve. The result of the Zn^{++} ions going into solution is to build up an *excess of electrons on the zinc plate.*
3. The H^+ ions are repelled over to the copper plate from which each one takes an electron, thus becoming a neutral hydrogen

atom. These atoms unite in pairs to form hydrogen gas (H_2) which bubbles off at the copper plate. This part of the process creates an increasingly large *shortage of electrons on the copper plate.*

4. The chemistry of the process is completed by the uniting of the Zn^{++} ions and the SO_4^{--} ions to form zinc sulfate ($ZnSO_4$).

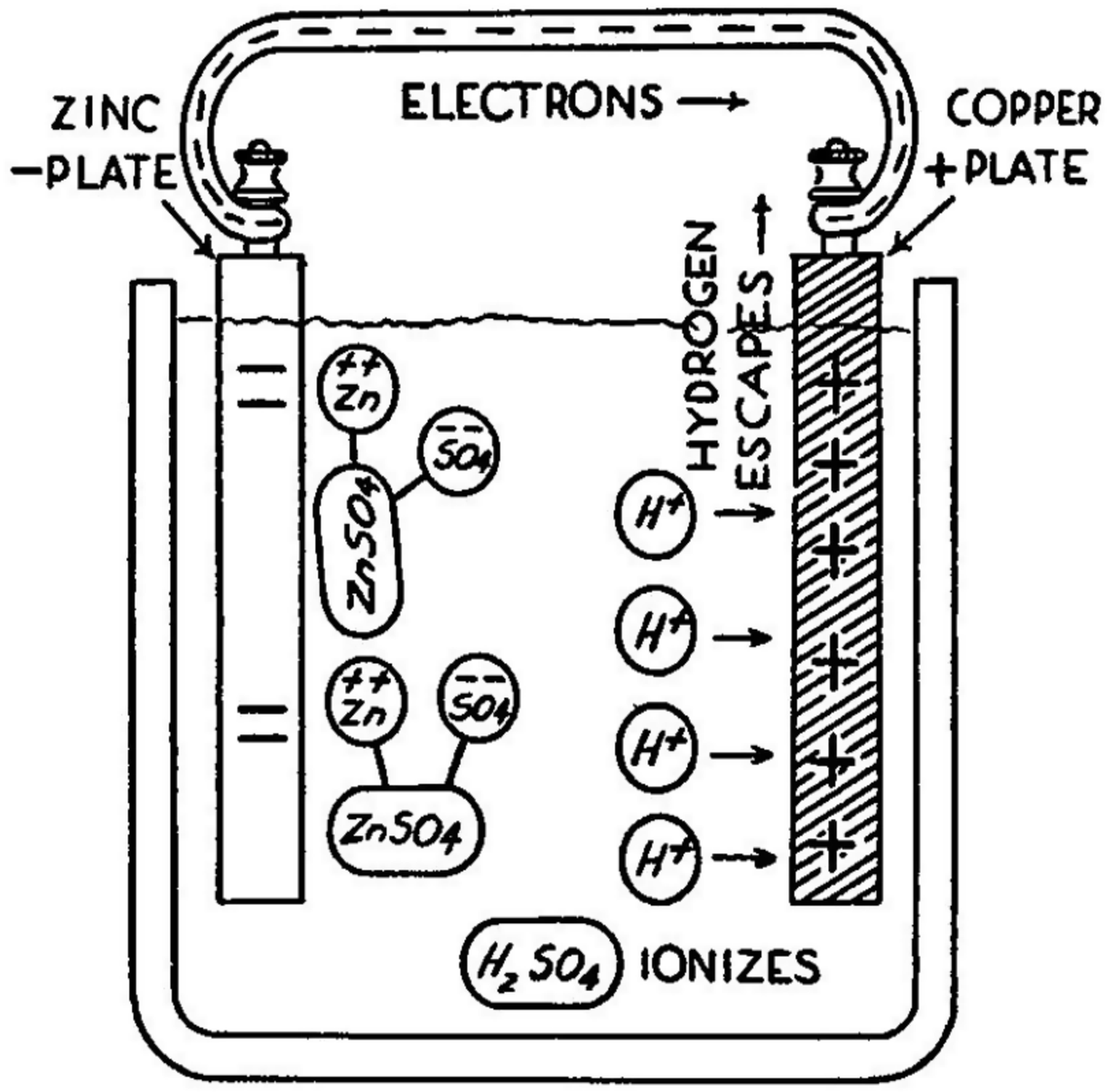


FIG. 140. Diagram illustrating action of a simple voltaic cell.

Thus, by means of chemical reactions, a set of conditions has been created such that one part of the cell, the zinc plate, possesses an excess of electrons over protons, while another part, the copper plate, has fewer electrons than protons. It is obvious that if these plates are connected by a conducting wire, forces of repulsion and attraction between charges will cause electrons to migrate from the zinc to the copper through the wire. In other words, an electric current will have been generated.

It will be noted that both the zinc and the sulfuric acid decrease in amount as the action proceeds. Thus the cell ceases to generate a current whenever one or the other of these is used up. Such a cell can only be made to work again by putting in new constituents. Cells of this type, that cannot be recharged, are called primary cells.

One point of possible confusion in the above description should be mentioned. Many years before the discovery of the electron and the fact that an electric current is a parade of electrons, it became de-

sirable for scientists to agree on a direction of flow for electricity. Since at that time there was no way to determine the facts about the flow, it was arbitrarily decided to picture it as flowing from positive to negative in the external circuit and it is this direction that is still used in most electrical circuit diagrams. But, in the action of the voltaic cell, we have just seen that the current actually consists of electrons flowing from the negative plate to the positive plate. Today we speak of both currents, calling the imaginary, arbitrarily chosen one the conventional current, and the actual one the electronic current.

Other Primary Cells. — A number of other primary cells or batteries have been developed with a view to overcoming certain undesirable features that are present in the simple one just described. The most familiar of these is the so-called dry cell, which really isn't dry at all. In it the hollow zinc cylinder which contains the other materials also serves as one of the plates or electrodes. The other electrode is a carbon rod in the center of the battery. The solution which dissociates is ammonium chloride (NH_4Cl) mixed with other things to form a moist paste. Names and descriptions of the action of various other primary cells can be found in elementary physics and chemistry textbooks.

In general, one can say that any two different conductors placed in a solution in which dissociation occurs will furnish an electric current, that is, will be an electric cell. The reader can test this by placing a copper cent and a silver dime on his tongue. If he lets the two coins touch each other, he will experience an unpleasant taste as a result of the electrical action that is set up. Incidentally, this action may be very undesirable. For instance, if gold and silver fillings touch each other in one's mouth, the two in conjunction with the saliva constitute a little electric cell that will dissolve out the silver. It should be mentioned that two metals in a solution of the sort described does not constitute the only way in which a primary cell can be made. It is possible to accomplish a similar result with one metal and two solutions or with two metals and two solutions.

Before entirely leaving the question of cells of this sort, the reader should turn back to page 299 in Chapter 17 and read again the discussion of the term " electromotive series " in relation to " chemical activitv series."

Storage Batteries. — Secondary or storage batteries are current-producing cells in which the chemical reactions are reversible. This means that, when the battery is "run down," it can be recharged by passing an electric current through it in the opposite direction to that of the flow when the battery is in use. The most common cell of this type is the lead storage battery used in automobiles. The plates in such a battery are lead grills, the negative plates being filled with spongy lead (Pb) and the positive with lead peroxide (PbO_2). They are placed in the container in such a manner that the two kinds alternate. The solution, or electrolyte, like that in the voltaic cell previously discussed, is sulfuric acid (H_2SO_4) in water. The chemical reactions that occur, both when the cell is being used and when it is being charged, are as follows:

Charged Condition		*Discharged Condition*			
Lead	+ Lead peroxide +	Sulfuric acid	⇄	Lead sulfate	+ Water
Pb	+ PbO_2 +	$2H_2SO_4$	⇄	$2PbSO_4$	+ $2H_2O$

During discharge, that is while being used, the reaction is from left to right. When the battery is being charged the reaction is from right to left. Most readers will be familiar with the fact that one usually tests the condition of an automobile battery with a hydrometer which actually measures the density of the solution. It will be evident from the above equation why the density of the liquid is a measure of the condition of the battery. When fully charged the cell contains a maximum of sulfuric acid which is a heavy liquid. During discharge the percentage of acid present decreases and that of water, a much lighter liquid, increases.

Lead storage cells have certain undesirable features among which are their great weight, the fact that they do not hold up well under repeated physical shocks, and their rapid deterioration when uncharged. These defects are largely absent from the Edison storage cell. In it the negative electrode is of iron oxide, the positive one of nickel hydroxide, and the liquid a solution of potassium hydroxide. The disadvantages of the Edison type battery are that it is somewhat more expensive than the lead battery and delivers a lower voltage.

Generation by Use of a Magnetic Field

Chemical methods of generating electric currents were the first to be discovered and were the only source of current for half a century or more. However, the enormous electrical development of the present day would have been quite impossible without the invention of the dynamo, a device which makes use of a relationship between electricity and magnetism in generating an electric current. In fact, thousands of separate chemical cells would be required to provide currents of the magnitude that are used today. Just to produce an electric arc between carbon poles, one experimenter in the year 1808 used 2,000 cells. The expense involved and the labor required to keep such outfits in working condition would prohibit their use.

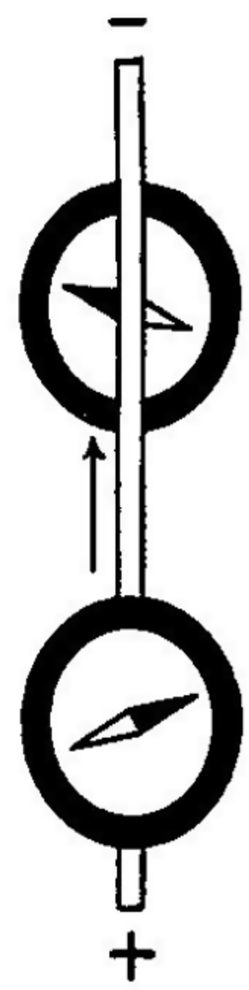


FIG. 141. Diagram showing how a compass needle is deflected above and below a current-carrying wire.

Historical Background. — We have suggested above that a dynamo uses magnetism to produce electricity. As a matter of fact the first actual proof of the existence of a relationship between the two came from an experiment which demonstrated the reverse effect, namely, that electricity produces magnetism. The Danish physicist, Hans Christian Oersted (1770–1851), one day in 1819, happened to place a compass under a current-carrying wire and was startled to see that the needle no longer pointed north but took up a position approximately at right angles to the wire, as shown by the upper compass in Fig. 141. The lower instrument indicates the manner in which the needle behaved when he placed it above the wire. In both cases the shaded end of the compass needle represents the north-seeking pole. It should also be noted in this diagram that the arrow indicates the direction of the " conventional current " as defined earlier in this chapter. Although the true current is the electronic one, we will use the conventional current in all of the diagrams in this section on magnetic methods of generating a current in order not to confuse readers who may already be familiar with certain hand rules for the determination of field and current directions.

By placing compass needles at other points near the current-carrying wire, Oersted demonstrated that the magnetic field about the wire was circular, as shown at the top of Fig. 142. The lower part of this figure illustrates one of the hand rules mentioned above. It will be noted that when the wire is grasped by the right hand with the thumb pointing in the direction of the conventional current, the fingers circle the wire in the direction of the magnetic lines of force.

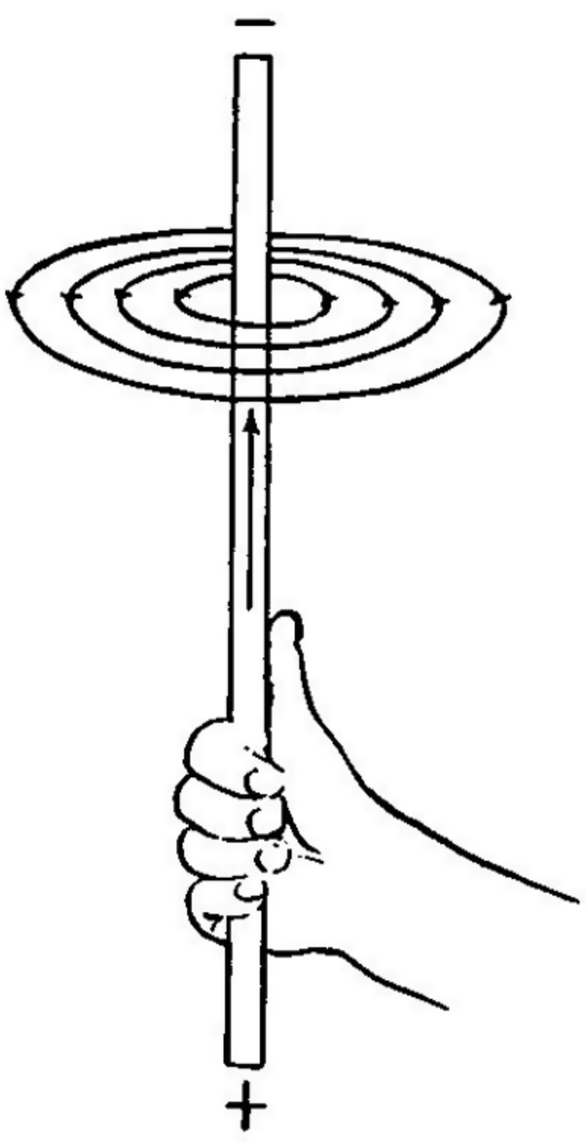


FIG. 142. Diagram showing the magnetic field about a current-carrying wire and the right-hand rule for determining its direction.

As soon as Oersted announced the above results a number of other scientists began to experiment in this field, and new discoveries followed quickly. Among them was the demonstration by the Frenchman, André Ampère (1775–1836), that a coil of wire which carries a current behaves like a magnet. The magnetic field about such a coil, or *solenoid*, is shown in (*a*) of Fig. 143. The strength of the magnetic field can be greatly increased by placing an iron core within the coil as in (*b*). When this is done the outfit becomes an electromagnet, one of the most useful modern devices based on the relationship between electricity and magnetism.

The discovery which laid the foundation for the development of the dynamo and therefore of our whole modern system of generating electrical energy was made by the English experimental physicist, Michael Faraday (1791–1867). Faraday was familiar with the work mentioned above, showing that a magnetic field resulted when an electric current flowed through a conductor. He became interested in the possibility of reversing this effect and obtaining an electric current from a magnetic field. After many unsuccessful attempts he discovered the experimental fact which solved the problem. That fact may be stated as follows:

Whenever an electrical conductor cuts through the lines of force in a magnetic field, an urge is created for electrons to flow along the wire.

This urge will be shown to be what we call electromotive force or voltage.

The Simple Dynamo. — Having seen something of the background of man's knowledge of the relationship between electricity and magnetism, we will now ask what sort of device based on these phenomena might be used as a source of current.

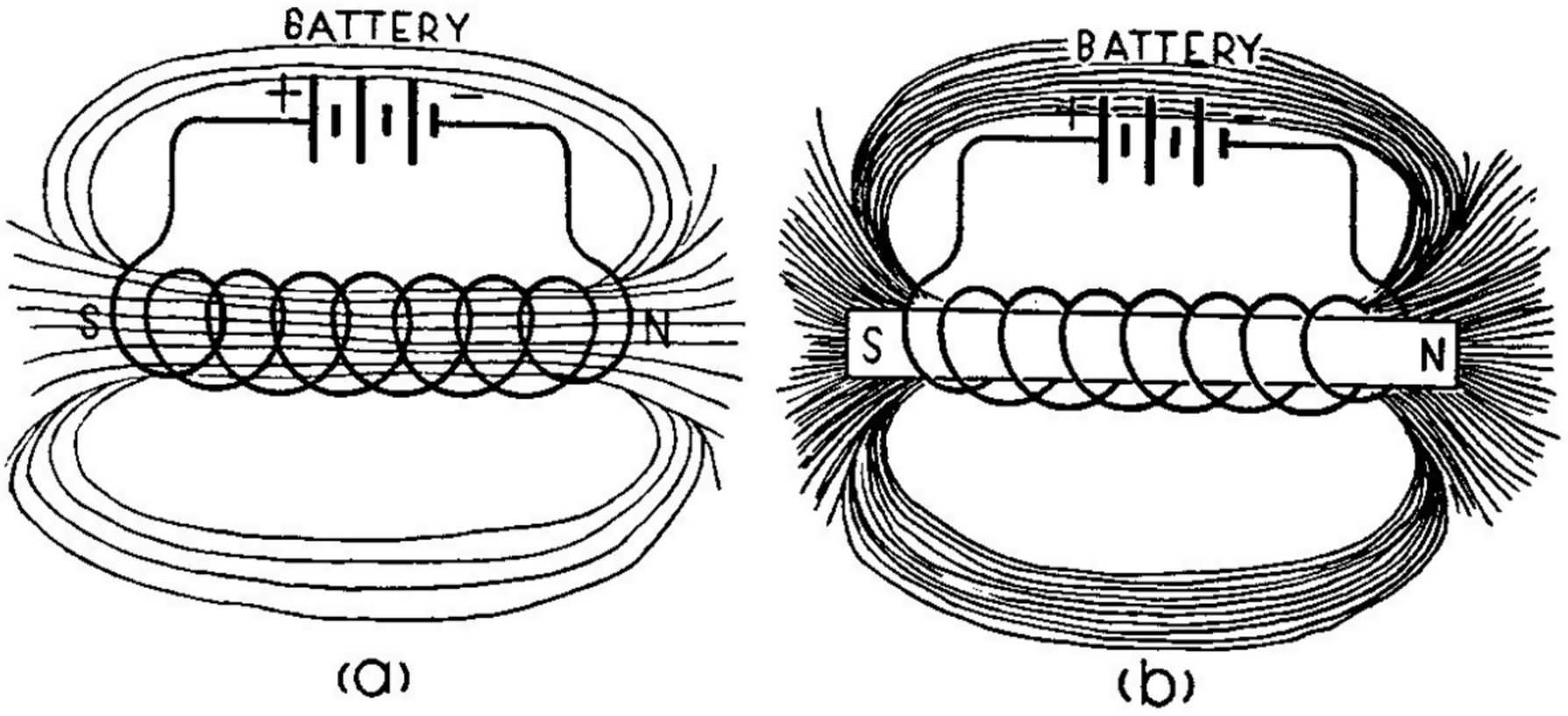


FIG. 143. Diagrams showing the magnetic fields about (a) a solenoid and (b) an electromagnet.

Perhaps the simplest form of dynamo would be a single wire made to move between two bar magnets, as shown in Fig. 144. The wire in moving down cuts through the magnetic lines of force, shown by the dotted lines, thus fulfilling the condition mentioned previously for the generation of an electric current. The mechanical energy that is converted into electric energy comes from the work that must be done to move the wire through the field. The lines of force resist this motion, somewhat as rubber bands might do, and so force must be exerted to move the wire. Lifting the wire up through the field causes the current to tend to flow in the opposite direction. A stronger field and consequently a larger current can be obtained by using electromagnets instead of bar magnets to produce the field.

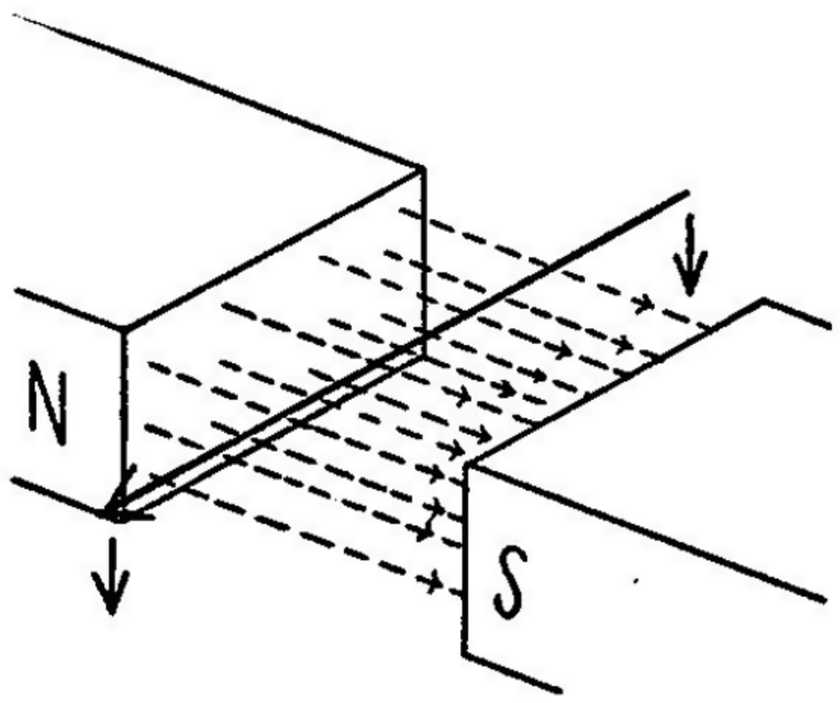


FIG. 144. Diagram illustrating the effect of moving a straight wire through a magnetic field.

The above outfit would not be very satisfactory as a source of continuous current. A better arrangement, and one that is really a

simple form of dynamo, can be obtained by using a rotating loop instead of a single wire. A generator of this sort is shown in Fig. 145. Note that electromagnets are used in producing the magnetic field, part of the current generated by the dynamo being diverted for use in this manner. Now as the loop rotates, both sides of it cut lines of force. However, since each side is moving up half of the time and down the other half, the current will change direction twice every revolution. A current which reverses in this fashion is called an *alternating current* and is the sort that is supplied by most power companies. It is commonly of a kind called " 60 cycle alternating current " which means there are 60 complete cycles, or 120 reversals, of the current each second. In order to conduct the current out from the dynamo it is necessary to use an arrangement of rings and brushes, also shown in the diagram. Each end of the loop is connected to a ring (R) which turns as the loop rotates. So-called brushes (BB) then rub on the rings collecting the electrical energy and conducting it to the outside circuit.

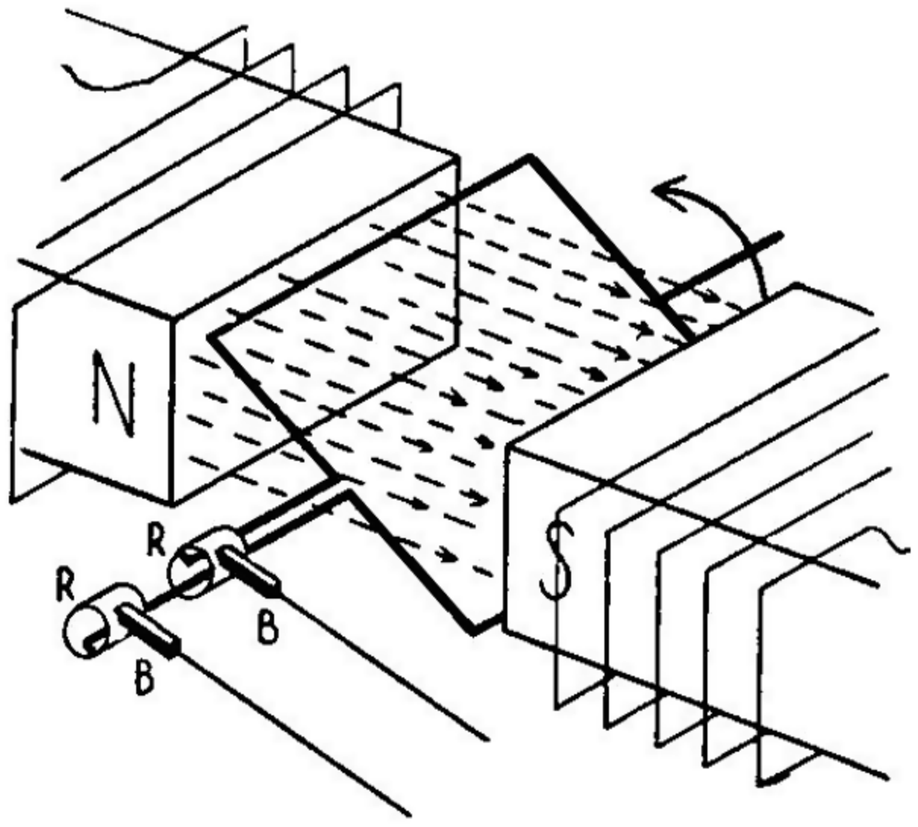


FIG. 145. Diagram illustrating the operation of a simple dynamo of generator. (After *Electricity and Wheels*, General Motors Corp.)

In the simple dynamo just described, all the essentials of a modern generator are present. Large industrial dynamos differ from it only in details. For one thing, the single loop is replaced by many loops, the entire rotating portion being called an armature. Multiplying in this fashion the number of wires that cut lines of force increases the size of the current and makes the flow more even. Also various types of winding are employed in making the electromagnets which furnish the field. If a direct current rather than an alternating one is desired, a device called a commutator is added. The commutator in no way affects the manner of generating the current, but simply interchanges the contacts that lead to the outside circuit every time the current in the armature reverses.

Some outside source of energy must be employed, of course, to

turn the armature in a dynamo. When it is available, water power is used by large power companies. Otherwise steam engines, turbines, or Diesel engines provide the motive power. Small electric power plants designed to furnish electricity for a home or farm are commonly run by gasoline engines. The dynamo or generator on an automobile, which furnishes electricity for the lights, starter, and ignition systems, is run by the engine itself.

Generation by the Action of Heat

This phenomenon, known as the *thermoelectric effect*, was described in Chapter 5 in connection with our description of an instrument for measuring surface temperatures of stars. The instrument is the thermocouple and we will refer the reader back to that discussion, which appears on page 84.

Generation by the Action of Light

When light falls on certain metals, including potassium, sodium, and zinc, electrons are emitted from the metal. This phenomenon, called the *photoelectric effect*, can be demonstrated as shown in Fig. 146. A highly polished zinc plate is connected to an electroscope and both are charged negatively. Now if light is allowed to fall on the zinc, the leaves of the electroscope will come together, indicating that electrons are escaping. That is, a flow of electricity has been set up as a result of the action of light. If the plate and electroscope are charged positively instead of negatively, the light has no effect.

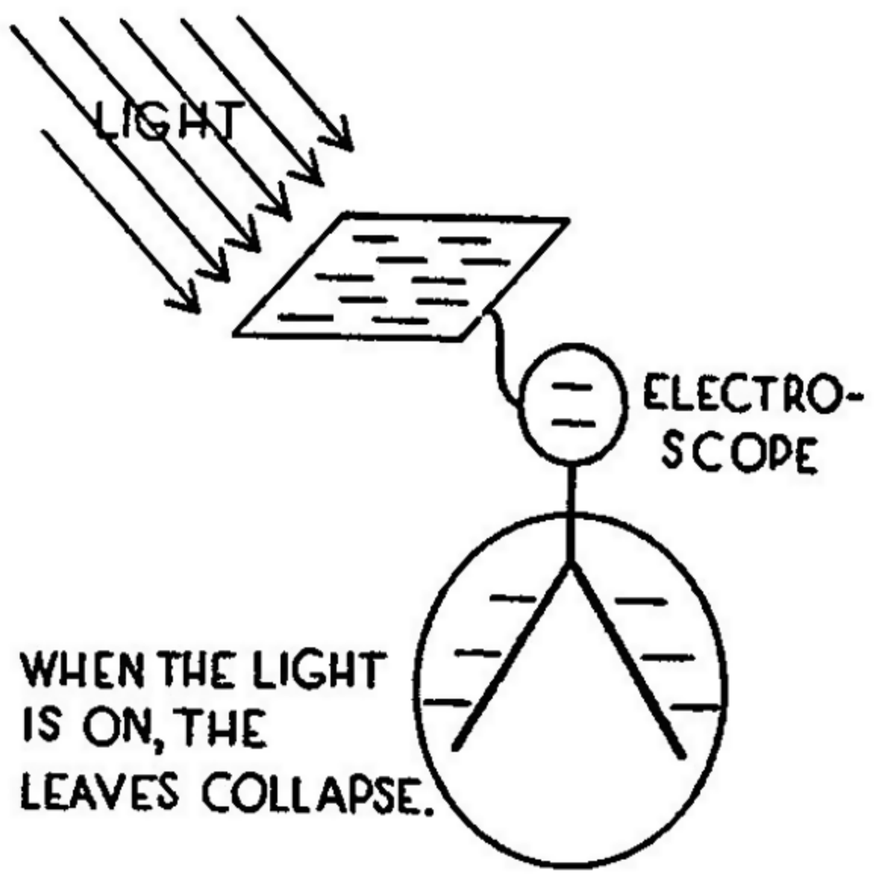


FIG. 146. Diagram illustrating the photoelectric effect.

A commercial device which makes use of this effect is the photoelectric cell or so-called "electric eye." It acts essentially as a switch that can be turned on or off by the action of a beam of light. Its practical applications today are numerous and varied. Automobiles on highways or going through tunnels are counted by having each one pass through a beam of light which shines on an electric eye. Similar

devices can be used to count the output of a factory or the number of people entering a building. Photoelectric burglar alarms utilize the invisible ultra-violet rays of light. The beam is so directed that an intruder will pass through it, cutting off the light, and giving the alarm. With an electric eye, automobile headlights can be used to operate a switch that opens the garage doors, street lights or school-room lights can be automatically turned on and off as the outdoor light varies in intensity, or a waitress with a loaded tray can open the kitchen door by walking through a beam of light. Also, it is the electric eye which makes sound motion pictures possible, since it enables sound to be recorded as varying shades of density on the film and then reproduced in the theater as sound.

The Electric Circuit

Although an exhaustive consideration of electrical units of measurement would be out of place in this book, there are a few that must be discussed briefly because they concern fundamental characteristics of a simple electric circuit. Their names, such as volt, ampere, ohm, and watt, are familiar to most people, but judging from their use in common conversation, their exact meanings are frequently somewhat obscure.

Analogy between an Electric Circuit and a Water Circuit

In many respects the flow of electricity through a conductor is analogous to the flow of water through a pipe. For one thing, in the latter case it is necessary to have a pressure exerted in order to force the water through the pipes to the consumer. This pressure is developed either by a pump or by having the liquid raised to a considerable height in a standpipe or reservoir. The pressure can be measured in pounds per square inch. In an electric circuit, the urge which sends the electrons through the conductors can be thought of as an electrical pressure. It happens to be called *electromotive force* and is usually designated simply as E.M.F. This electrical pressure is measured in *volts*. Voltages of the sources of electricity discussed above vary greatly. For example, the E.M.F. developed by a thermocouple may be only a few millionths of a volt while large commercial generators produce voltages up in the thousands. It will be recalled that the

maximum voltage developed by the Van de Graaff electrostatic generator, previously discussed, was in the millions. A dry cell has an E.M.F. of 1.5 volts while a single lead storage cell develops a maximum of 2 volts. The ordinary 6-volt automobile storage battery consists of 3 single lead cells. Such a 6-volt battery will send four times as many electrons through a given circuit in a certain length of time as would be forced through by a 1.5 volt dry cell, just as a pressure of 60 pounds per square inch will force four times as much water through a given pipe as would one of 15 pounds per square inch in the same length of time.

The quantity of water that goes through a pipe can be measured in gallons or cubic feet. We have already defined the fundamental unit of quantity of electricity as the charge that repels a like charge, one centimeter distant in air, with a force of one dyne. Since this unit is very small, a more practical one, called the *coulomb* is used. A coulomb is 3×10^9 times as large as the smaller unit.

The rate at which water flows through a pipe might be measured in gallons per minute or in cubic feet per minute. The rate of flow of electrical energy is measured in *amperes*. If one coulomb of electricity flows through the circuit every second, then the current is said to be one ampere.

The pipes through which water runs offer a certain amount of resistance to its flow. Although there is no specific unit of water resistance, it might be measured in terms of the force of friction between the water and the pipe. Three things on which such friction would depend occur to one immediately. It seems obvious that the larger the pipe the less the resistance it would offer, other things being equal. On the other hand, we would expect resistance to be directly proportional to the pipe's length. The third factor would be the roughness of the inner surface. Electrical conductors also offer resistance to the flow of electricity. In this case there is a common unit, one that is called the *ohm*. Two factors governing electrical resistance are exactly analogous to the first two mentioned for water. Resistance decreases as the diameter of the conductor gets larger, and is directly proportional to the length of the conducting wire. A third factor, in the case of electrical resistance, is the material out of which the conductor is made. Silver is the best conductor of electricity, which means that

it offers the least resistance. Copper and aluminum also have very low resistances, which makes them valuable for use in transmission lines. Substances like rubber, glass and other non-metals have such high resistances that we call them non-conductors or insulators.

The three electrical quantities of E.M.F., current, and resistance are related by an experimental law known as Ohm's law. It is perhaps the most important relation in the entire field of electricity and may be stated as follows:

$$Current = \frac{E.M.F.}{Resistance}$$

or

$$Amperes = \frac{Volts}{Ohms}$$

The comparison between units in the two kinds of circuits, as described above, is summarized in Table 22.

TABLE 22. COMPARISON OF UNITS IN WATER AND ELECTRICAL CIRCUITS

Property to be Measured	*Units Used in Water Circuit*	*Units Used in Electric Circuit*
Pressure	Pounds per square inch	Volts
Quantity	Gallons	Coulombs
Rate of flow	Gallons per minute	Amperes
Resistance	(none)	Ohms

Power in Electric Circuits

In Chapter 20, power was defined as rate of doing work. Therefore, in an electrical circuit, power may be thought of as the rate at which electrical energy is being used. The commonly used electrical units of power are the watt and kilowatt. The relation of power in watts to the units discussed above is given by,

$$Volts \times Amperes = Watts.$$

Thus, if a 6-volt storage battery is delivering a current of 2 amperes, electrical energy is being expended at a rate of 6×2, or 12 watts. The power rating of electric light bulbs is stamped on them as 60W meaning 60 watts.

Now let us see how electrical power is related to the electric bill that arrives each month. This was mentioned briefly in the former

discussion of the watt referred to above. A kilowatt is equal to 1,000 watts. A kilowatt-hour (KWH) is a unit of electrical work or energy and is the amount of energy that would be used by a 1,000 watt bulb or stove if it burned for 1 hour. Obviously, 1 KWH of electricity would also be used by a 500 watt bulb in 2 hours, or a 100 watt one in 10 hours. In other words,

$$Kilowatt\text{-}Hours = Kilowatts \times Hours.$$

It is this kilowatt-hour that is the unit used in determining the electric bill. The rate is usually a sliding one, the cost per KWH decreasing as the total number used increases.

Applications of Electricity

Many of the uses that we make of electricity are well known to everyone. It is our purpose here simply to summarize and classify these applications and mention a few of the less familiar features about them.

Heating Effect of an Electric Current

When an electric current passes through a wire, heat is generated and the temperature of the wire is raised. The amount of heat developed in a given time is proportional to the resistance of the wire and to the square of the current. Therefore, in applications where the formation of heat is desired, wires of high resistance are used. Perhaps the most important such application is the ordinary electric light bulb. All such incandescent sources emit light because the filament is raised to a very high temperature. The thin tungsten filament in such a lamp has a resistance equal to that of more than 20 miles of ordinary lamp cord. Incidentally, because of the large proportion of energy that goes into heat, incandescent lamps are very inefficient as light sources, a matter that will be discussed in some detail in Chapter 33.

Other household appliances which make use of this heating effect are toasters, cooking ranges, water heaters, flatirons, bathroom heaters, and waffle irons. In all of them wire of high resistance is used. When a fuse burns out it does so because of the heating effect of an electric current. A 15-ampere fuse, for example, contains a metal alloy element that will melt and break the circuit if a current of more than 15

amperes goes through it. In this way it protects the building from possible disastrous results due to accidental overloads on the line. Another application, the electric furnace used in the manufacture of high grade steel was mentioned in Chapter 17.

In this section we have been discussing applications in which it is desirable that an electric current should cause heating. When such heating is undesirable the current should be kept as small as possible, because of the fact that the heat generated increases with the square of the current.

Mechanical Effect of an Electric Current

The electric motor, which is a device that converts electrical energy into mechanical energy, is based on an experimental fact essentially the converse of the one given for the dynamo. It may be stated,

If an electric current is made to flow through a conductor that is in a magnetic field, there will be exerted on the conductor a force at right angles both to the field and to the direction of the current.

The essential similarity of construction of dynamos and motors is shown in Fig. 147. In the dynamo, the loop, or armature, is turned mechanically by an engine, waterfall, windmill or turbine. As it rotates, magnetic lines of force are cut and a current is generated. This current is conducted through an outside circuit to a motor. In the motor, the loop is stationary in the magnetic field as long as no current flows through it. But as soon as electrons begin to move through the wire there is a force set up which causes the loop to rotate, developing mechanical energy. By attaching a belt to the pulley on the motor shaft, this mechanical energy can be employed to run machines of various sorts. The dynamo and motor shown here are both of the direct current type. An alternating current motor is based on the same principle but is constructed somewhat differently. It will be evident from the construction as described above that either of the devices could be used as a dynamo or as a motor. This is also true of commercial motors and generators of this sort. The operation of some electric-powered trains offers an example of a motor doubling for a dynamo. In mountainous regions, the motor uses electrical energy from the line in pulling the train up the mountain. Then, coast-

ing down the other side under the influence of gravity, the train runs the device as a generator, putting electrical energy back into the line. It is a little as though one's automobile while coasting down one hill could be made to operate the engine in such a manner as to manufacture and pump back into the tank part of the gasoline necessary to take it up the next hill.

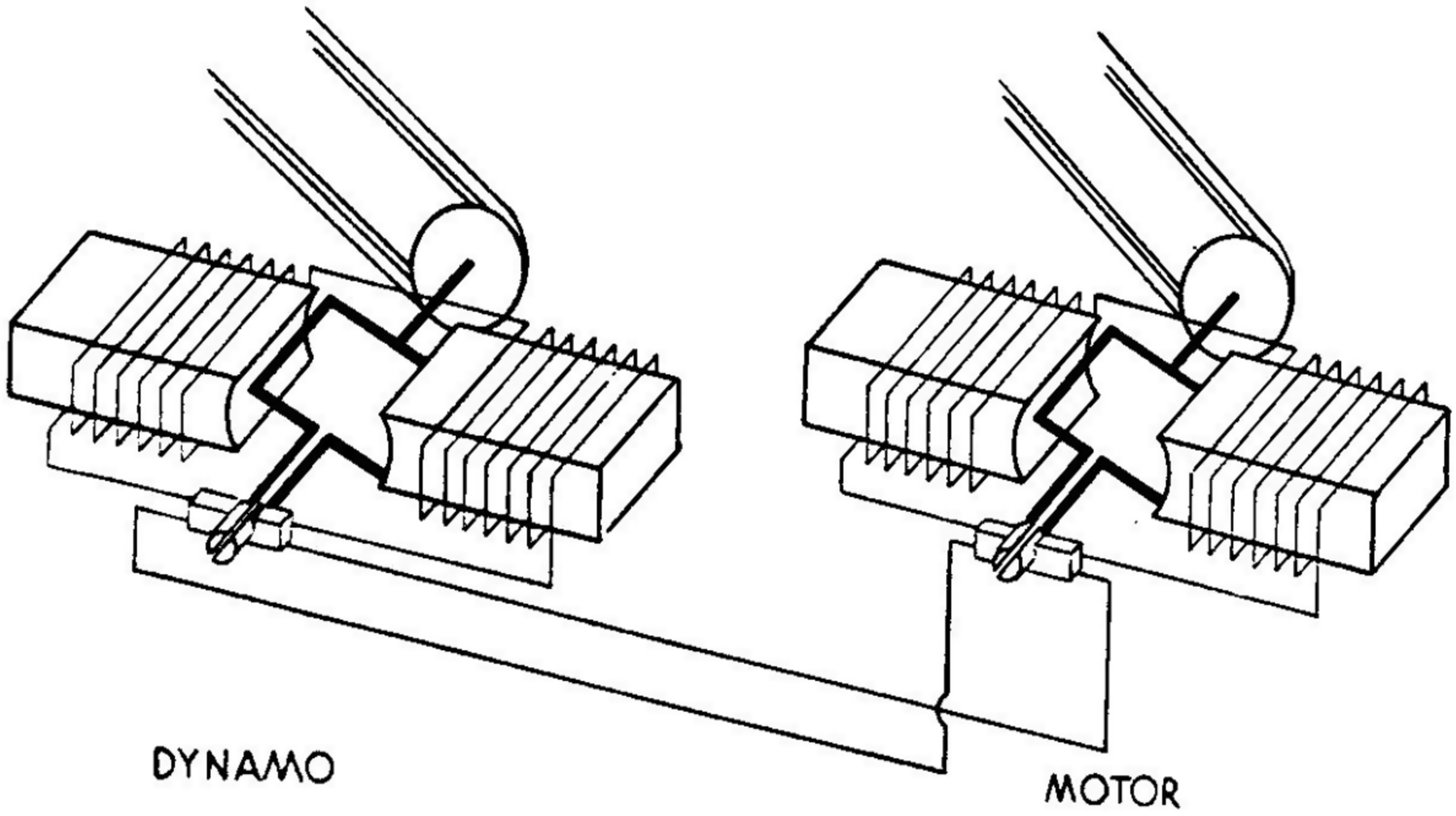


FIG. 147. Diagram illustrating operation of direct-current dynamo and motor. (After *Electricity and Wheels*, General Motors Corp.)

The energy transformations accomplished by dynamos and motors, together with the fact that electrical energy can be transmitted over considerable distances fairly efficiently, make it possible for man to use natural energy sources such as waterfalls over a large area. Suppose energy could only be transmitted from one place to another mechanically, that is, by pulleys, belts, shafts, or gears. It would be quite impractical to attempt to use the energy of Niagara Falls to operate a machine shop even as close as Buffalo. But electrically, it is both practical and profitable to convert the mechanical energy of the falling water into electrical energy by means of a dynamo at Niagara Falls, then to conduct this electrical energy over transmission lines to Buffalo, and finally at Buffalo to use a motor to change it back into mechanical energy to run the machine shop.

Chemical Effects of an Electric Current

Electrolysis of Water. — We are now in a position to consider in some detail the decomposition of water by the passage of an electric current through it, a process to which reference has been made several

times previously in this book. An experimental set-up for electrolysis of water is shown in Fig. 148. Since it is not desirable that the outfit should act as a battery, the plates are made of the same metal, preferably an inactive one like platinum. A battery or other source of direct current is connected to the plates making, in this case, the left-hand

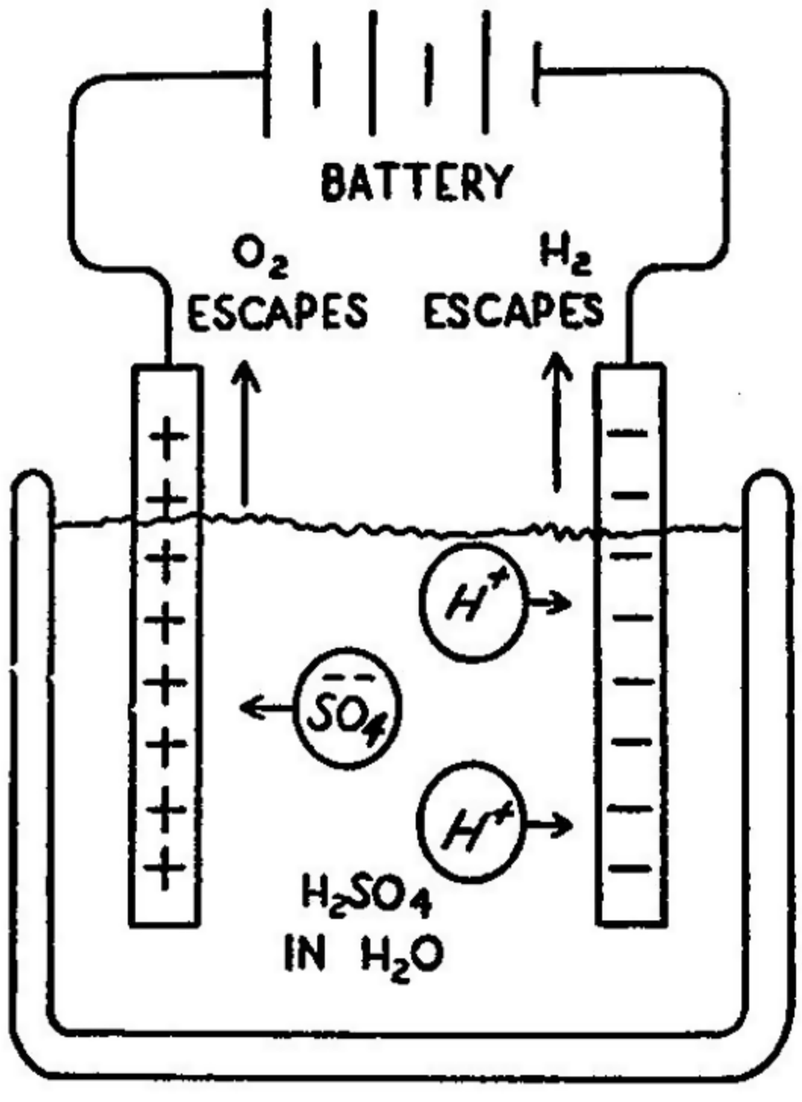


FIG. 148. Diagram illustrating electrolysis of water.

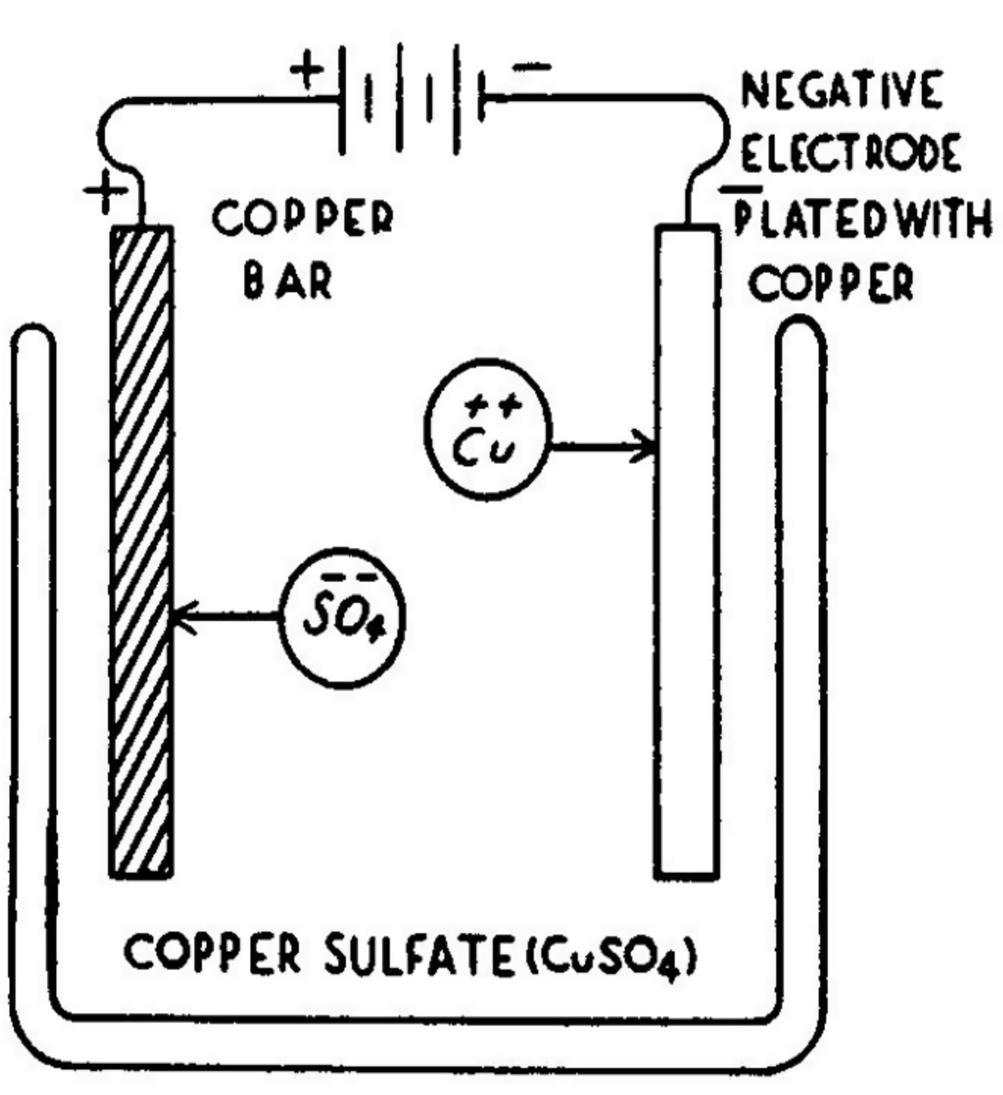


FIG. 149. Diagram illustrating the process of copper-plating.

one positive (*shortage of electrons*) and the right-hand one negative (*excess of electrons*). The container is filled with water in which there is some sulfuric acid. The steps in the electrolytic process may be stated as follows:

1. The sulfuric acid dissociates, as previously explained.
2. The H^+ ions are attracted to the right-hand plate where each one acquires an electron and becomes a neutral hydrogen atom.
3. The neutral hydrogen atoms group into pairs and escape at the negative plate as diatomic hydrogen gas (H_2).
4. The SO_4^{--} ions are attracted to the left-hand plate where they give up their excess electrons.
5. Each neutral SO_4 group displaces the oxygen atom from a molecule of water (H_2O) and appropriates the H_2 part for itself, thus forming a new molecule of sulfuric acid (H_2SO_4).
6. The oxygen atoms, thus displaced, group into pairs and escape at the positive plate as diatomic oxygen gas (O_2).

It will be noted that the amount of sulfuric acid in the container remains constant, since a new molecule is formed (Step 5) for every one that dissociates (Step 1). Thus water is the only material that is permanently decomposed and the net result of the reaction can be expressed by the equation that we have used previously, namely,

$$2H_2O \rightarrow 2H_2 + O_2$$

Other Examples of Electrolytic Action. — The experimental set-up pictured in Fig. 149 differs from the preceding one in that the positive platinum plate has been replaced by a copper plate, and a solution of copper sulfate ($CuSO_4$) substituted for the sulfuric acid in water. The copper sulfate ionizes forming Cu^{++} and SO_4^{--} ions. When an electric current is sent through the solution, the SO_4^{--} ions are attracted over to the positive copper bar where each one is neutralized and combines with an atom of copper to form another molecule of copper sulfate. The Cu^{++} ions, on the other hand, are attracted to the negative plate, become neutral, and are deposited on the plate as pure copper. Thus, as the action continues, the positive copper bar dissolves and an even layer of copper is built up on whatever object serves as the negative electrode. This process, which is called *electroplating*, has numerous practical applications.

The printing of this book offers one illustration of its use. After each page was set up in regular type-metal, a mold of it was made in a certain kind of wax. This wax impression was then coated with graphite to make its surface an electrical conductor and was placed in an electrolytic cell as the negative electrode. A layer of copper was deposited on it, in the manner described above, reproducing the printing surface of the type. When the copper layer reached the desired thickness it was removed from the wax and reinforced with a type-metal backing. This particular application of electroplating is called electrotyping.

The plating process can be used with a number of metals. In every case the object to be plated constitutes the negative electrode; a bar of the plating metal is the positive electrode; and the electrolyte is a solution of the plating metal. Other examples include plating silver on spoons and other articles, gold on watches and jewelry, chromium on various metal objects, and zinc on steel to prevent rusting.

Three previous references have been made in this book to practical applications of this sort of process. Two of them concerned the metallurgy of copper and aluminum, respectively, discussed in Chapter 17. In the case of copper the method is a refining one. A bar of impure copper is used as the positive electrode in an electrolytic cell. The solution is copper sulfate. Since the impurities are not carried across to the negative electrode during electrolysis the deposit built up on it is the pure metal. In the metallurgy of aluminum, it will be recalled, electrolysis is the reduction process. The aluminum ore is dissolved in molten cryolite and a current passed through the solution. During the process the aluminum separates out and collects as a liquid at the bottom of the container. The third previous reference was in Chapter 18 in the description of ways of covering objects with a coating of rubber. Since the colloidal particles of latex carry a negative charge they can be attracted by a positive charge. If an electric current is passed through a solution of latex, the rubber particles will be deposited or " plated " on the positive electrode.

The Transmission of Electrical Energy

It is usually not desirable to transmit electrical energy at the same current and voltage as that at which it is generated. One of the principal reasons for this is that an electrical current produces in a conductor, a heating effect which is proportional to the square of the current. Other things being equal, a 2-ampere current will produce 2^2 or 4 times the heating loss of a 1-ampere current, or a 4-ampere current 16 times that of a current of 1 ampere. Now it will be recalled that electrical power may be expressed by,

$$Watts = Volts \times Amperes.$$

This indicates that if, in a given case, one were to increase or " step up " the voltage the current would be decreased proportionately, the power remaining the same. For example, a current of 3 amperes at 100 volts would have the same wattage as one of 1 ampere at 300 volts. The heating loss, however, would be 3^2 or 9 times as great in the former case as in the latter. In transmitting electrical energy over considerable distances it is desirable to step the voltage up in order to cut down

line losses. At the point of consumption then, it can be stepped back down again and the current increased.

The device which makes possible these changes in current and voltage is the *transformer*. Two experimental facts that have already been discussed have a bearing on the operation of a transformer. The first of these is the observation that any wire which is carrying an electrical current has a magnetic field about it. The second is the basis of operation of a dynamo, namely, that whenever a conductor cuts through magnetic lines of force a voltage is induced in that conductor. Now with these two facts in mind let us turn to Fig. 150 which is a diagram of a simple transformer. The ends of the coil of wire marked "primary" are connected to a source of alternating current such as a dynamo. Consequently, the coil will have a magnetic field about it all the time that current is flowing in it. However, every time the current reverses direction there will be an instant when no current flows in the primary which means an instant during which there is no magnetic field about it. Thus, we can think of the field about the primary as alternately building up and collapsing with the reversals of the alternating current. Next, note the coil labelled "secondary." Every time the magnetic field about the primary builds up it is cut by the secondary and every time it collapses it is cut again. Consequently, in line with the second experimental fact mentioned above, an alternating current is induced in the secondary coil. With respect to the voltage relationships it happens to be true that the ratio between the voltage of the secondary and that of the primary is the same as the ratio between the number of turns of wire on the two. For example, if a transformer has 100 turns on the primary and 10,000 on the secondary, then a primary voltage of 110 volts would be stepped up to 100 times 110, or 11,000 volts on the secondary. The current would be correspondingly reduced. Such a transformer is called *step-up* transformer since the voltage is increased. One that has fewer

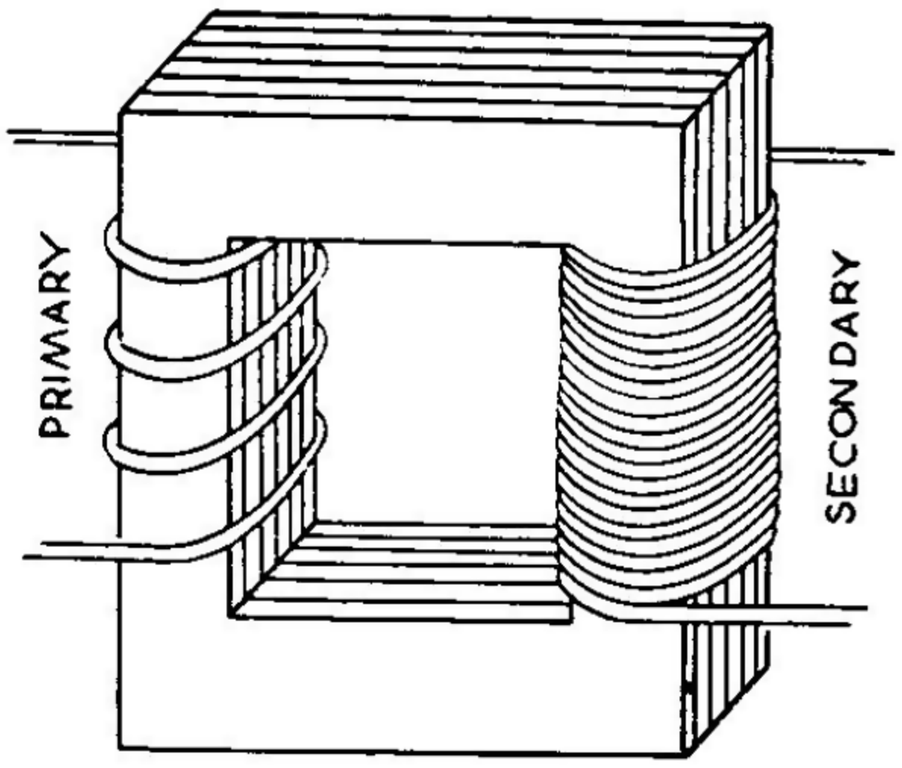


FIG. 150. Diagram showing the primary and secondary coils in a simple transformer.

turns of wire on the secondary than on the primary is called a *step-down* transformer, since the induced secondary voltage is less than that on the primary coil. Wrapping the coils on an iron core, as shown in the figure, increases the strength of the magnetic fields. It should be noted that there is no direct electrical connection between the primary and secondary.

Typical voltages employed in the transmission of electrical energy are the following. The generator at the power plant develops 13,200 volts. This is stepped up to from 60,000 to 260,000 volts for transmission to another city for the reason mentioned above. Then on the outskirts of the city it is stepped down to 13,200 volts and at a few places in the city to 2200 or 2300. Then near the place of consumption further reduction occurs to the usual 110, 115, or 220 volts.

Transformers find many applications in addition to those directly connected with long distance transmission of electrical energy. They are useful in any situation where it is desirable to either increase or decrease the voltage of an alternating current.

Questions

Discussion

1. What is the difference between a static discharge and an electric current?
2. Look up one accomplishment each of Galvani and Volta, other than the ones mentioned in this chapter.
3. From your knowledge of the chemical activity series, decide which set of discs in Volta's voltaic pile were the negative electrodes.
4. Work out the various steps in the action of a voltaic cell made of zinc and copper electrodes but using hydrochloric acid (HCl) in water as the electrolyte.
5. What is the difference between a hydrogen atom and a hydrogen ion?
6. Which of the constituents remain constant in amount during the action of each of the following? Explain.
   a. Electrolysis of water.
   b. Voltaic cell.
   c. Copper plating.
7. Distinguish between conventional current and electronic current.
8. What are the bare essentials for
   a. Chemical generation of an electric current?
   b. Magnetic generation of an electric current?

c. Thermoelectric generation of an electric current?

d. Photoelectric generation of an electric current?

9. What does a hydrometer measure and why can it be used to check the condition of a lead storage battery?

10. Draw diagrams to show what direction the current would be flowing in a north and south wire if a compass needle,

a. Points east when placed under the wire;

b. Points west when placed above the wire.

11. Observe what the brushes look like and where they are located on an actual generator or motor.

12. Mention two reasons for using electromagnets rather than bar magnets to excite the magnetic field in a dynamo.

13. Find as many actual practical applications of electric eyes and thermocouples as you can.

14. If the resistance of a circuit remains constant, how will,

a. Doubling the voltage affect the current?

b. Halving the voltage affect the power?

15. Find out what your local electric rates are and then verify your last month's electric bill.

16. Look up the relationship between kilowatts and horsepower in an earlier chapter and then calculate how many foot-pounds of work are done by 10 cents worth of electrical energy at your local rates.

17. What is the difference between the energy conversion accomplished by a generator and that by a motor?

18. Look up the details of some specific electroplating operation other than the examples mentioned in this chapter.

19. Would you expect to be able to use a transformer with direct current? Explain.

Multiple Choice

1. Which one of the following electrical units concerns a quantity in an electric circuit analogous to that measured in a water circuit by the ordinary water meter? (a) Volt. (b) Ohm. (c) Ampere. (d) Coulomb. (e) Watt.

2. During both the electrolysis of water and the operation of a simple voltaic cell, (a) the amount of sulfuric acid present remains constant, (b) one of the electrodes dissolves, (c) hydrogen escapes at the negative electrode, (d) oxygen escapes at the positive electrode, (e) the amount of water present remains constant.

3. At 4 cents per KWH, the weekly cost of operating two 100-watt lamps, six 60-watt lamps, a 100-watt radio and a 1-kilowatt heater an average of 3 hours per day would be between (a) 50¢ and $1; (b) $1 and $1.50; (c) $1.50 and $3; (d) $3 and $5; (e) $5 and $10.

4. List all of the following that are present in a fully charged automo bile storage battery. (a) A zinc container. (b) A carbon electrode. (c More lead sulfate than lead. (d) Sulfuric acid. (e) A compound of lead an oxygen.

5. In electroplating an object with copper (a) the solution used shoul be sulfuric acid, (b) electrons from the source of current enter the platin cell at the copper electrode, (c) the copper is plated on the positive electrode (d) hydrogen is given off at the negative electrode, (e) the positive electrod should be made of copper.

True–False

1. Electrolysis of water works best if the two electrodes are of differen metals rather than of the same one.

2. Luigi Galvani was a physician rather than a physicist.

3. Volta's voltaic pile was actually a whole series of individual electri cells.

4. Most of the electrical energy used commercially today is generated b chemical reaction.

5. During the operation of the simple voltaic cell, the zinc plate gradu ally becomes coated with copper.

6. A cubic inch of the liquid in a fully charged lead storage batter weighs more than a cubic inch of the liquid in the same battery when it i partially discharged.

7. Five 50-watt light bulbs burning for four hours would use just on kilowatt-hour of electrical energy.

8. Something that could be measured in kilowatt-hours could also b measured in foot-pounds.

9. If a lamp with a resistance of 15 ohms is connected to a 60-volt battery 4 coulombs of electrical energy will flow through the lamp every second.

10. In an electrical circuit, watts times ohms divided by volts equals volts

Chapter 26

ELECTRICITY AND THE ATOM

It will be recalled that in beginning our discussion of the nature of matter, in Chapter 13, we said the story would be told in two episodes. The part having to do with the atomic composition of bulk matter was covered at that time. We are now ready, to the extent that it is possible in this book, to consider the makeup of the atom itself.

Before looking at experimental evidence that throws light on the inside of the atom, it may be well to summarize briefly the picture of matter as we left it at the close of Chapter 14.

Review of the Picture of Matter as Previously Discussed

The kinetic theory was presented in the earlier discussion as offering an interpretation of the nature of matter and heat that is in agreement with experimental facts. The basic assumptions of that theory are that all matter is composed of molecules and atoms, as those terms were defined; that the molecules and atoms are in constant motion, with velocities that increase with temperature; and that they exert certain forces of attraction on each other. At the close of that part of the discussion we were left with some 92 building blocks of matter, that is, 92 particles out of which the multitude of objects we see about us is made. We are now about to examine the composition of these 92 building blocks and to consider a group of experimental facts which, on investigation, would seem to indicate that the atoms are, in turn, combinations of a much smaller number of basic bits of matter. This possibility has already been implied in our earlier reference to the electron and proton in Chapters 24 and 25. These must now be fitted into the atomic picture.

Some Experimental Facts Which Throw Light on the Nature of the Atom

The phenomena we are about to consider are among the discoveries which have completely revolutionized man's concept of matter during the past 50 years. They have served to open up a vast new field of atomic research out of which has come, among other things, the knowledge and skill that made possible the development of the atomic bomb which brought World War II to its spectacular close. It is perhaps worth noting that the first of these discoveries came shortly after several prominent physicists had stated that nothing remained to be done in physics but to measure certain constants to more decimal places.

It should be emphasized once more in this connection that because of mathematical limitations our discussion of these phenomena and what they mean to the story of the atom must be largely qualitative in nature and must therefore carry with it the danger of oversimplification. As far as possible, however, we will attempt to point out the limitations of the methods used and the extent to which conclusions follow directly from experiment.

Flow of Electricity through Gases

It is well known that gases, under normal conditions, are poor conductors of electricity. However, by reducing its pressure, air or any other gas can be made to conduct electricity. It was the study of this phenomenon that laid the groundwork for the discovery of the electron.

The Effect of Pressure on Conductivity. — The name of Heinrich Geissler (1814–1879), a German glass blower, is associated with the early devices used to demonstrate the fact that electricity will flow through gases which are at low pressure. A modern demonstration outfit, very similar in construction to the tubes used by Geissler in some of his experiments, is shown in Fig. 151. Metal electrodes are sealed into the ends of a glass tube which, preferably, should be several feet long. A side tube leads to a vacuum pump. Now, suppose an E.M.F. of 10,000 or 15,000 volts is connected across the tube and the vacuum pump started. At first there is no visible effect, but

as the pressure is reduced, long bluish streamers of light gradually appear, extending from one electrode to the other. Still further evacuation causes these streamers to broaden out until they fill the entire tube. By this time a reasonably large current is flowing, indicating that the electrical resistance of the gas has been greatly reduced. The familiar neon sign exemplifies one application of such a device. The bearing of light sources of this kind on the whole question of illumination will be considered in Chapter 33. Still further evacua-

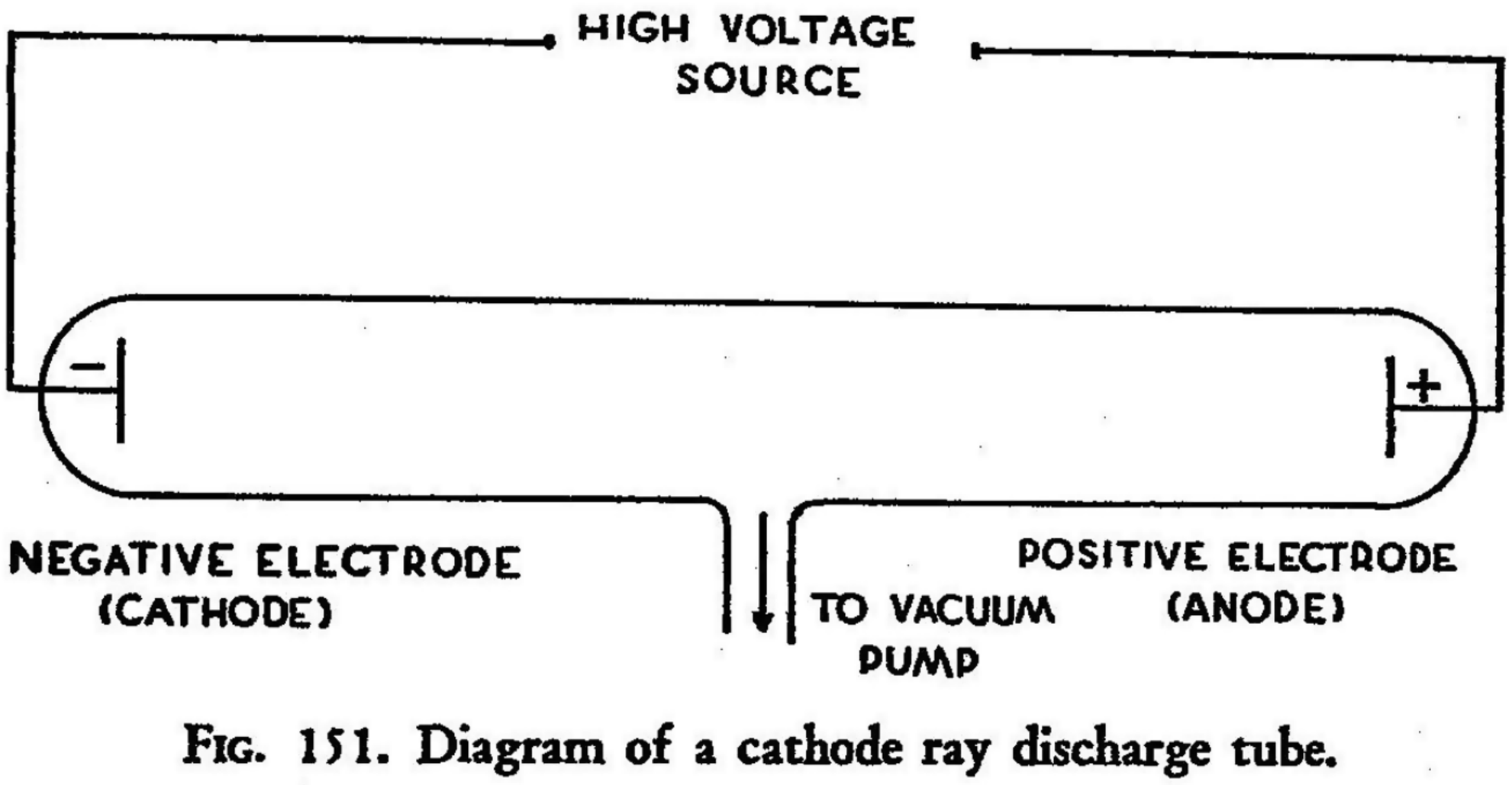


FIG. 151. Diagram of a cathode ray discharge tube.

tion is marked by an increase in resistance and the appearance around the cathode (negative electrode) of a dark space which expands and finally, at very low pressure, fills the entire tube. Although light is not being emitted by the gas, the glass walls of the tube itself now glow with a kind of greenish light.

Cathode Rays. — The "dark space and greenish glow" stage of the phenomenon, mentioned at the close of the preceding paragraph, was discovered by the famous English scientist, Sir William Crookes (1832–1919). The greenish glow proved to be a fluorescence of the glass which Crookes believed was caused by invisible radiation of some sort shooting out of the cathode. These radiations he called *cathode rays*. Using discharge tubes of the shape pictured in Fig. 152, Crookes set about investigating the properties of the radiations in order to be able to draw some conclusion as to their nature. The experimental procedures used by Crookes and others and the arguments based on them were as follows:

1. The tube shown in Fig. 152 demonstrated that the *cathode rays travel in straight lines*, by reason of the fact that the object fixed inside cast a sharp shadow on the end of the tube. This effect, it was obvious, might be produced either by radiation of the nature of light or by particles.
2. When a concave cathode was substituted for a flat one, *the rays came to a focus*. A tiny piece of platinum located at the point

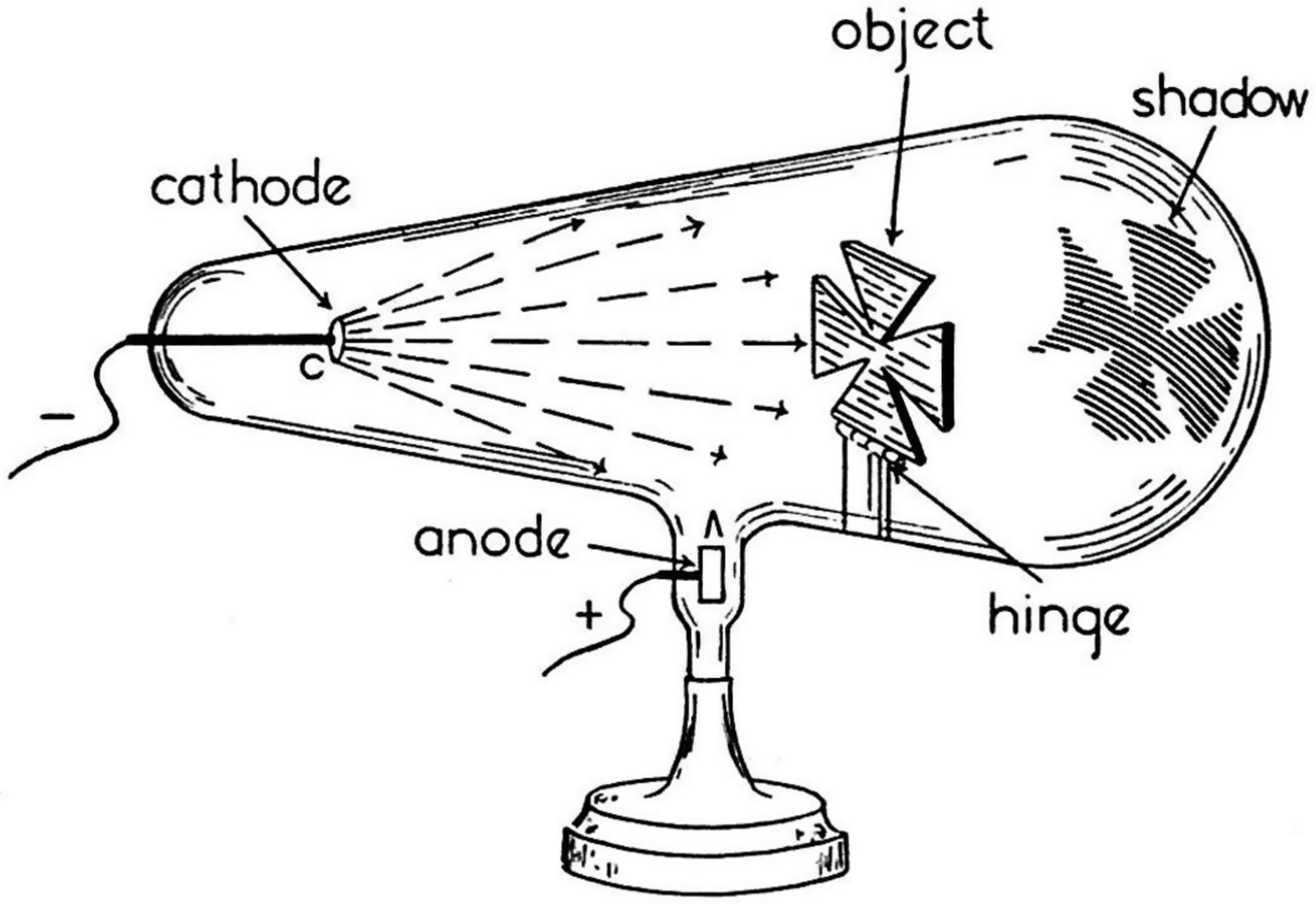


FIG. 152. A Crookes tube. (After White.)

of focus soon became red hot, indicating that the cathode radiation had energy. Again the result was one that might be produced either by light-like rays or by particles.
3. When the maltese cross in the Crookes tube was replaced by a small paddle-wheel, so fixed that the rays could strike only the top vanes, *the cathode rays caused the paddle-wheel to rotate*. This seemed to suggest that the radiation consisted of tiny bullets rather than light waves.
4. When the tube was so arranged that the rays passed either between electrically charged plates or magnetic poles, it was found that the fluorescent spot on the end of the tube moved up or down, depending on the direction of the field. Thus, *the cathode rays were deflected by electric and magnetic fields*. In the electrical case, the cathode-ray stream moved toward the positive plate and away from the negative plate, which seemed

to indicate not only that the radiation consisted of particles but that the particles were negatively charged. This conclusion was substantiated by the manner of deflection in the magnetic field in which the cathode rays moved at right angles to the direction of the field in accordance with the condition we have already stated for the operation of an electric motor.

The above paragraphs summarize the experiments and conclusions which, by 1895, constituted man's knowledge of the flow of electricity through gases. It should be noted that the results of research up to that time were entirely qualitative in nature, giving no information about the masses of the particles, or the magnitude of the electric charges they carried; nor was there any indication whether or not all of these particles were alike.

Discovery and Investigation of the Electron

Measurements Made by Sir J. J. Thomson. — The logical method of determining whether the cathode-ray particles were all alike was to attempt to measure their charge and mass. The results of investigations carried out along this line by the English physicist, Sir J. J. Thomson (1856–1940), were such that the discovery of the electron is commonly credited to him. The particular experiment conducted by Thomson which indicated all the cathode particles to be alike marked the actual beginning of atomic research and since the technique he employed has become classical in the study of such phenomena, we will consider its qualitative aspects in some detail. In that experiment Thomson used a modified form of Crookes tube, the shape and construction of which is shown in Fig. 153. The shields *DD*, with tiny holes at their centers, served to limit the cathode stream to a thin pencil of rays, which struck the fluorescent screen at S. Plates *PP*, which could be given opposite electric charges, were so arranged that the cathode stream passed between them. Charged as shown in the diagram, they would serve to bend the rays up, making them strike the screen at some point such as *M*. The larger the charge on the plates, the greater would be the amount of bending. The coils *MM* served as electromagnets, so that by properly passing a current through them, a north-seeking pole could be obtained in front of the

tube and a south-seeking pole behind it. With such an arrangement the cathode rays would bend down, striking the screen at some point *N*. Here also, the amount of the deflection could be varied, this time by varying the current in the coils.

The experiment was performed in two steps. In the first part, an electric charge was put on the plates deflecting the cathode rays one way, and then a magnetic field was applied of just sufficient strength to bring the rays back to their normal position. In other

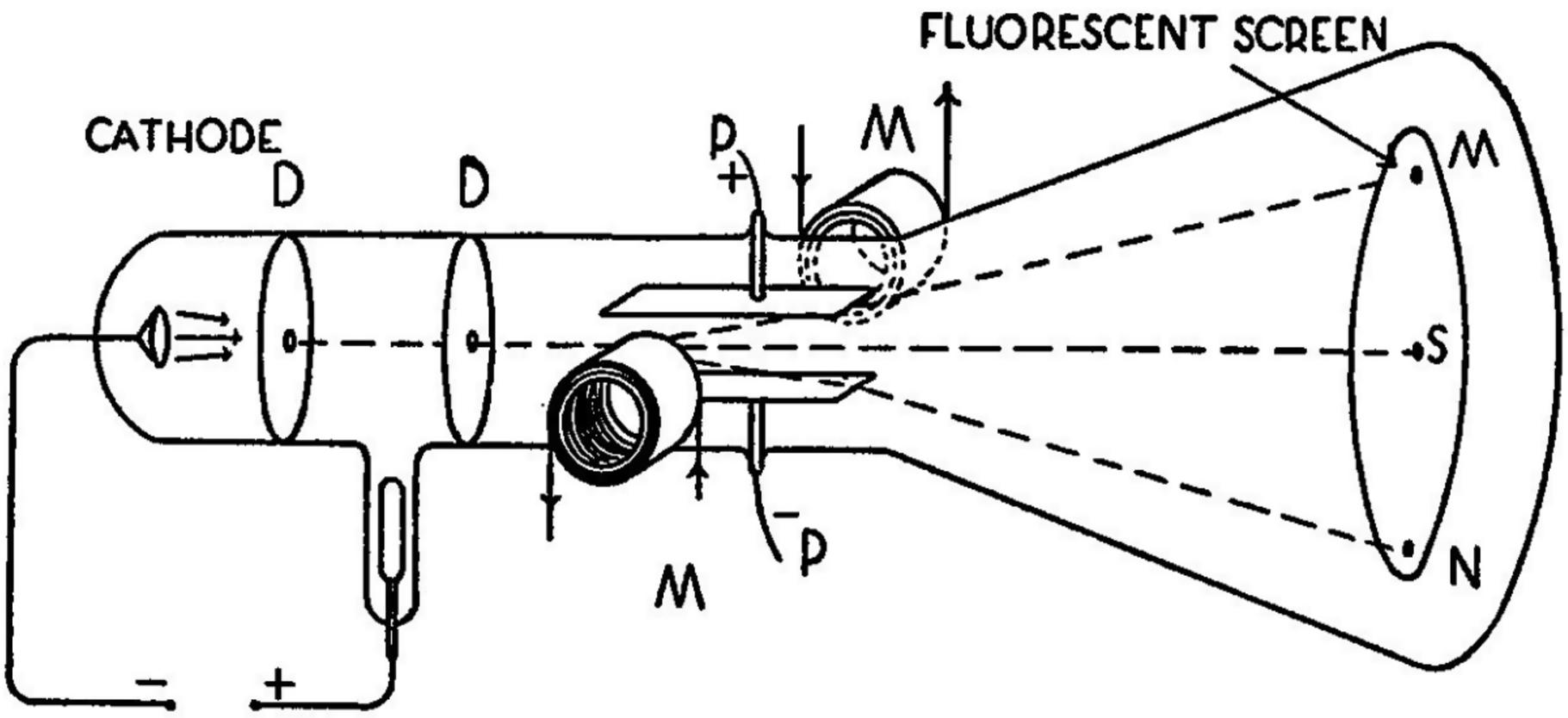


FIG. 153. Diagram of type of discharge tube used by Sir J. J. Thomson in measuring *e/m* of the electron. (After White.)

words, the magnetic and electric fields were made to balance each other exactly. The data taken were simply the strength of the two fields. In the second part, only a magnetic field was applied. The amount that the rays were deflected and the strength of the field were noted. From these two sets of data, together with the dimension of the apparatus, Thomson was able to develop a mathematical equation that contained two unknown quantities. These were the two things he wished to determine, namely, the mass of a cathode particle and the electric charge on one.

Here we will digress for a very brief discussion of equations which contain two unknown quantities. The fact about such equations that concerns us here is that while it is not possible to calculate the value of either unknown quantity by itself, it is possible to determine the ratio between them.

For example, suppose all we know about the prices of light bulbs

and mouse traps is that three of the former cost the same as two of the latter. This fact can be written in equation form as,

$$3 \times (\textit{cost of one bulb}) = 2 \times (\textit{cost of one trap}).$$

This equation contains two unknown quantities, the cost of one bulb and the cost of one trap. From this relation alone, neither of these unknowns can be calculated, since any number of values would fit. For instance, the equation would hold true for 10-cent bulbs and 15-cent traps, or 15-cent bulbs and 22½-cent traps, or 20-cent bulbs and 30-cent traps. We can, however, determine the ratio between the prices of the two articles. We can say that

$$\frac{\textit{Cost of one bulb}}{\textit{Cost of one trap}} = \frac{2}{3}$$

or, that a light bulb costs two-thirds as much as a mouse trap.

The equation developed by Thomson was exactly of the form described above. That is, it could be written (Some number) times (Charge on particle) = (Some number) times (Mass of particle) or, since the symbols commonly used for charge and mass in this connection are e and m, respectively,

$$(\textit{Some number}) \times e = (\textit{Some number}) \times m.$$

The parts of the equation which we have called "some number" were combinations of the actual measurements that Thomson made and have already been mentioned. From his equation, although he could calculate neither e nor m individually, he could determine a value for the ratio e/m, just as we have found one for the ratio between light bulb and mouse trap prices.

Now let us recall the original purpose of the experiment: One of Thomson's desires was to determine whether the particles in the cathode-ray stream were all alike. The nature of the above experiment makes it apparent that he really determined an average e/m for all the particles involved. He did not single out one particle and determine its e/m and then repeat for another, and so on, and then see if all the answers were the same. The reasoning which leads us to conclude from his work that the particles were all alike is this: Suppose there were all masses of particles mixed up together to form

cathode rays, and that they carried all sorts of electric charges. Then if one performed the Thomson experiment many times, one would expect that in some cases the average e/m would be different from that obtained in others. But if the particles were all alike the average e/m would always be the same. The latter case represents the experimental fact. The same experiment has been carried out thousands of times by Thomson and others, using tubes of various sizes and all sorts of magnetic and electric field strengths, and always the same value for the ratio e/m is obtained. Thus, before either one of these quantities had been measured separately, scientists felt justified in concluding that cathode-ray particles were all alike and constituted a fundamental bit of matter. It is these cathode rays that were first called *electrons*. As we shall see, since their discovery several other sources of electrons have been found.

Measurement Made by Robert A. Millikan. — The first accurate determination of the actual charge on the electron was made by the American physicist, Robert A. Millikan (1868–), for which achievement he was awarded a Nobel Prize in physics. The essential parts of his apparatus are shown schematically in Fig. 154. Two plates were attached to the terminals of a high voltage battery, and so arranged that the magnitude of the charge on them could be varied by the experimenter. The upper plate had a small hole in it which connected the space between the plates to the region above. The atomizer above the upper plate was used to produce a spray of oil drops. The X-ray tube was for the purpose of ionizing the gases between the plates. In this case the ionizing process was one of separating electrons from gas atoms. The positive and negative charges thus created then drifted about in the air. The experimental procedure was as follows:

The observer at the microscope would fix his attention on a single oil drop that fell through the hole in the upper plate. First, measurement was made of its rate of fall when there was no charge on the plates. From this its mass could be calculated. Then a charge was put on the plates, as shown in the diagram. The same oil drop was kept under observation for hours at a time. As it moved through the ionized air between the plates, it frequently acquired a negative

charge. Such an occurrence would cause it to start up toward the positive plate. Each time that this happened, the observer varied the charge on the plates enough to bring the oil drop back to rest. If it lost part of its negative charge it would start down, and the observer would have to adjust the plate charge in the opposite direction. The principal data consisted of a large number of adjustments of the electric charge on the plates, from each one of which the size of the

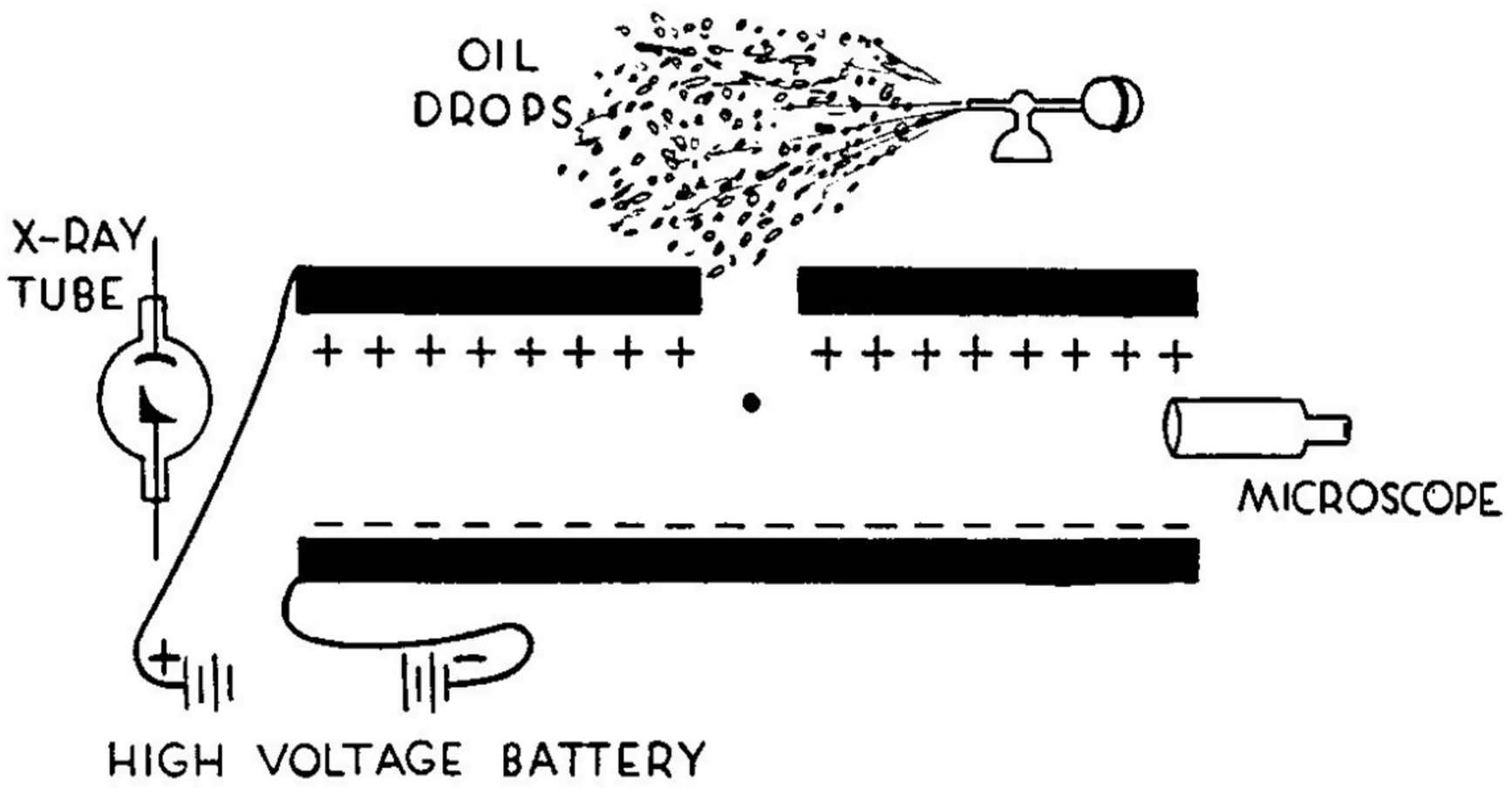


FIG. 154. Essential parts of the apparatus used by Dr. R. A. Millikan in measuring the charge of the electron. (After White.)

charge acquired or lost by the oil drop at one particular time could be calculated. Two predictions could be made about these hundreds of charges acquired by the oil drop. First, if there actually was a fundamental basic unit, or "atom," as it were, of electricity, then every one of the hundreds of charges picked up and lost by the oil drop should be a *whole multiple* of that unit. Second, the greatest common divisor of all the charges should be that unit, that is, the charge on the electron. Let us illustrate this reasoning by an example. Suppose that all one knew about the price of tickets at a certain motion picture theater was that the first five persons in line when the box office opened spent the following amounts of money:

1st individual —	$2.10
2nd individual —	3.50
3rd individual —	.70
4th individual —	1.75
5th individual —	1.05

From even these few data one could conclude that the price of one ticket had to be 35 cents, 7 cents, 5 cents, or 1 cent, since these are the only numbers that will divide an exact whole number of times into all five amounts. If several hundred people were investigated and it was found that no one spent less than 35 cents, and that all the amounts were divisible by 35, then one would feel safe in deciding, first, that there was a single definite price being charged for each ticket and, second, that 35 cents was that price.

Similarly, Millikan and many other investigators who performed the experiment have found that the charge acquired by the oil drop is never less than a certain amount and that it is always an exact whole multiple of that minimum value. It is this smallest charge that is known as e, the charge on the electron. It is so small an amount of electricity that it has to be multiplied by about 1.6×10^{19} in order to equal one coulomb.

The Mass of the Electron. — Once e/m and e are known, the mass m of the electron can be easily calculated; just as in our illustration, as soon as the price of a light bulb is known the cost of a mouse trap can be determined. Dividing e by e/m gives a value for the electron's mass of

$$m = .000,000,000,000,000,000,000,000,9 \text{ gram}$$
$$= 9 \times 10^{-28} \text{ grams},$$

the figure that was mentioned in Chapter 24.

We have covered these experiments on the electron in some detail for two reasons. First, they were classical investigations in atomic research and more or less served to initiate the enormous amount of work that has been done in that field during the past 25 or 30 years. Second, as far as this particular book is concerned, since we cannot describe all the experiments we will mention as fully as we have these, it is important that the reader should see these experiments as typifying the kind of indirect approach that must be employed in the investigation of atomic phenomena. Here we have considered a case in which neither the charge on the electron nor its mass, nor even proof of its existence, could be determined by direct observation. Nevertheless, experimental procedures were devised which led to relationships wherein the desired quantities were expressed in terms of

things that could be observed directly. This made possible the calculation of the electronic charge and mass. It should be emphasized that if such indirect procedures are well founded, they lead to results that are no less reliable than those obtained by methods which seem more direct. Incidentally, the more philosophically inclined reader will recognize that one might experience considerable difficulty in defining just where direct experimental procedures leave off and indirect methods begin.

Photoelectric Effect

This phenomenon was discussed in the preceding chapter as one of the methods of generating an electric current. We mention it again here simply to point out that it provides added evidence of the existence and nature of electrons. When the negative electricity emitted by metal plates in the photoelectric effect is studied by the Thomson and Millikan techniques, the e and e/m values obtained are the same as those determined for cathode rays.

Radioactivity

The Discovery of Radioactivity. — The discovery, in 1896, of the phenomenon of radioactivity and the subsequent determination of its nature provided undeniable evidence of the existence of particles of matter smaller than atoms. Radioactivity was discovered more or less accidentally by the French physicist, Henri Becquerel (1852–1908), while he was engaged in experiments with X-rays. He found that photographic plates, although wrapped in heavy black paper, would become exposed if they were near the element uranium or any of its compounds. This startling discovery excited considerable scientific interest and led to extensive research. By far the most important contributions to our knowledge of the phenomenon resulted from the work of Madame Marie Curie (1867–1934), whose name is familiar to everyone, and her husband, Pierre Curie (1859–1906). Her research in this field led to her sharing in two Nobel prize awards, one in physics and one in chemistry. Shortly after the discovery of radioactivity by Becquerel, the Curies set about looking for substances that would possess the property to a greater extent than did uranium. Their search led them to an ore called pitchblende, mined in Bohemia,

which seemed to be powerfully radioactive. After carrying out analyses of several tons of the ore, tiny amounts of three elements were isolated. The first one to be found was named *polonium*, for Madame Curie's native country, Poland. The other two were called *radium* and *actinium*. Of the three, only radium exists in quantities sufficiently large for chemical treatment. Some idea of the care and precision with which the Curies worked can be obtained from the fact that a ton of the most productive pitchblende yields only some two or three grains of radium. With the discovery of elements other than uranium that exhibited radioactive properties, the name given the radiation was changed from "uranium rays" to "Becquerel rays." Today a number of radioactive materials are known, all of them among the heavy elements.

The Nature of Radioactivity. — It might be desirable at this point for the reader to look back at the very brief and overly simplified description of radioactivity given in Chapter 9 in connection with the discussion of its use in determining the age of the earth. The point was made there that atoms of radioactive substances have the interesting habit of spontaneously shooting out parts of themselves as a result of which they become atoms of entirely different elements. The rays they emit, that is, those that Becquerel found affecting his photographic plates, are of three kinds. This can be demonstrated by the apparatus pictured in Fig. 155. A sample of radium, or one of its compounds, is placed at the bottom of a hole drilled in a block of lead. Thus, any radiation emitted by the radium must be moving vertically upward when it emerges from the block. Plates that can be charged as shown are located on either side of the block. With no charge on the plates a photographic film held above the apparatus will exhibit only one exposed spot, directly above the hole in the lead. But when the plates are given opposite electric charges the beam splits into three

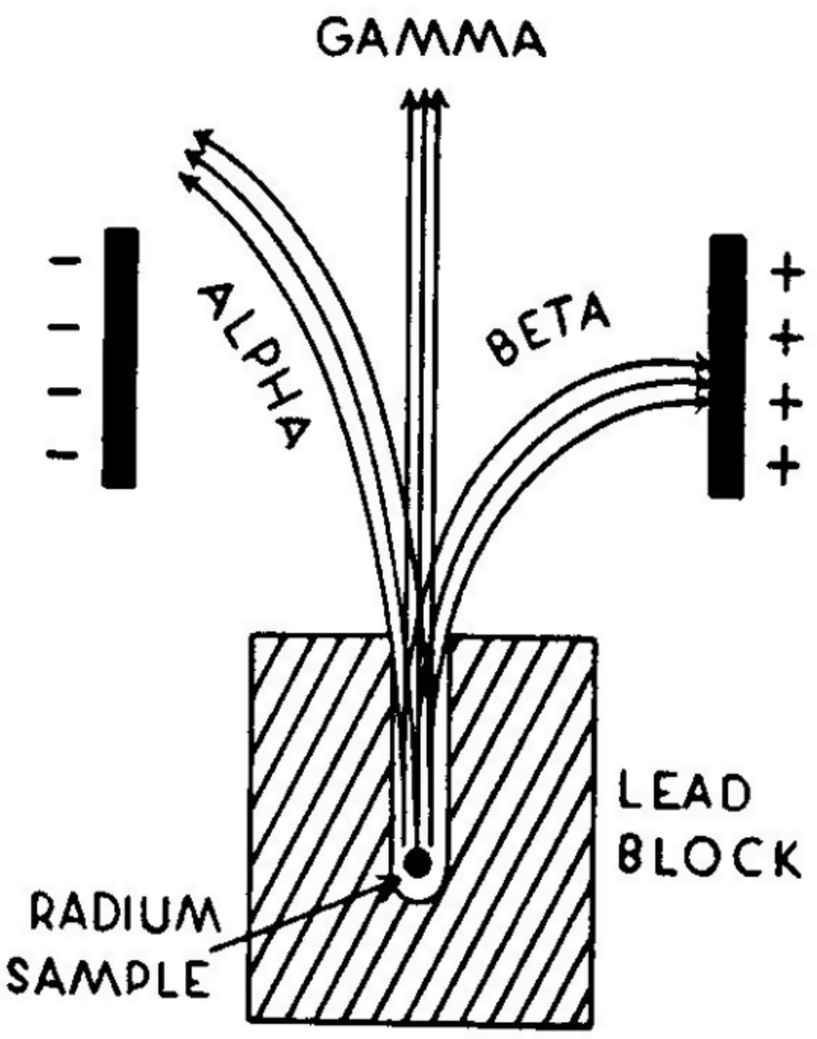


FIG. 155. Diagram illustrating the effect of an electric field on Becquerel rays.

parts as shown. Before their nature had been established the rays that were deflected toward the negative plate were called *alpha rays*, those that bent toward the positive plate *beta rays*, and the undeflected ones *gamma rays*. Additional experiments of a quantitative nature have established the following facts about the three types of radiation:

1. Alpha rays are particles each one of which has a mass about four times that of a hydrogen atom. In other words, the atomic weight of an alpha particle is 4. It has a net electric charge of +2. We shall see later that an alpha particle is identical with an atom of helium that has had two electrons removed from it. Alpha rays travel at velocities of from about one-hundredth to one-tenth that of light, that is, from 2,000 to 20,000 miles per second.
2. Beta rays are electrons, being identical in nature with cathode rays and photoelectrons. Emitted from radioactive substances they travel very rapidly, reaching in some cases a velocity 99 per cent that of light.
3. Gamma rays have the nature of light waves, differing from them only in length, the gamma waves being very small. Thus, they are not charged particles and so are not affected by electric and magnetic fields. They travel with the velocity of light. Specific characteristics and practical applications of gamma rays will be mentioned in Chapter 31.

There are several ways of detecting the presence and passage of Becquerel rays in addition to the demonstration pictured above. One of these, using what is known as a Wilson cloud chamber, makes it possible actually to see the paths of the alpha and beta rays. The cloud chamber consists of a device in which a small volume of air can be supersaturated with water vapor. Under such conditions, the vapor will condense, forming fog particles about any dust particles or ions that are present in the region. In this case, dust is kept out of the chamber to eliminate the first possibility. Now alpha and beta particles, as they plow through the air after being emitted from a radioactive atom, produce a train of ions, because of collisions with gas molecules. Thus, when they dart through the cloud chamber, their paths are marked by a fog streak as water vapor condenses on these

ions. The photographs in Fig. 156 show such cloud chamber paths for alpha and beta rays. The former, being much the heavier of the two, are less easily turned aside by collisions and so produce relatively straight paths. Beta rays, on the other hand, are frequently deflected. The ionization produced by gamma rays is too slight to be observed in this fashion. It will be obvious from the properties of the three types

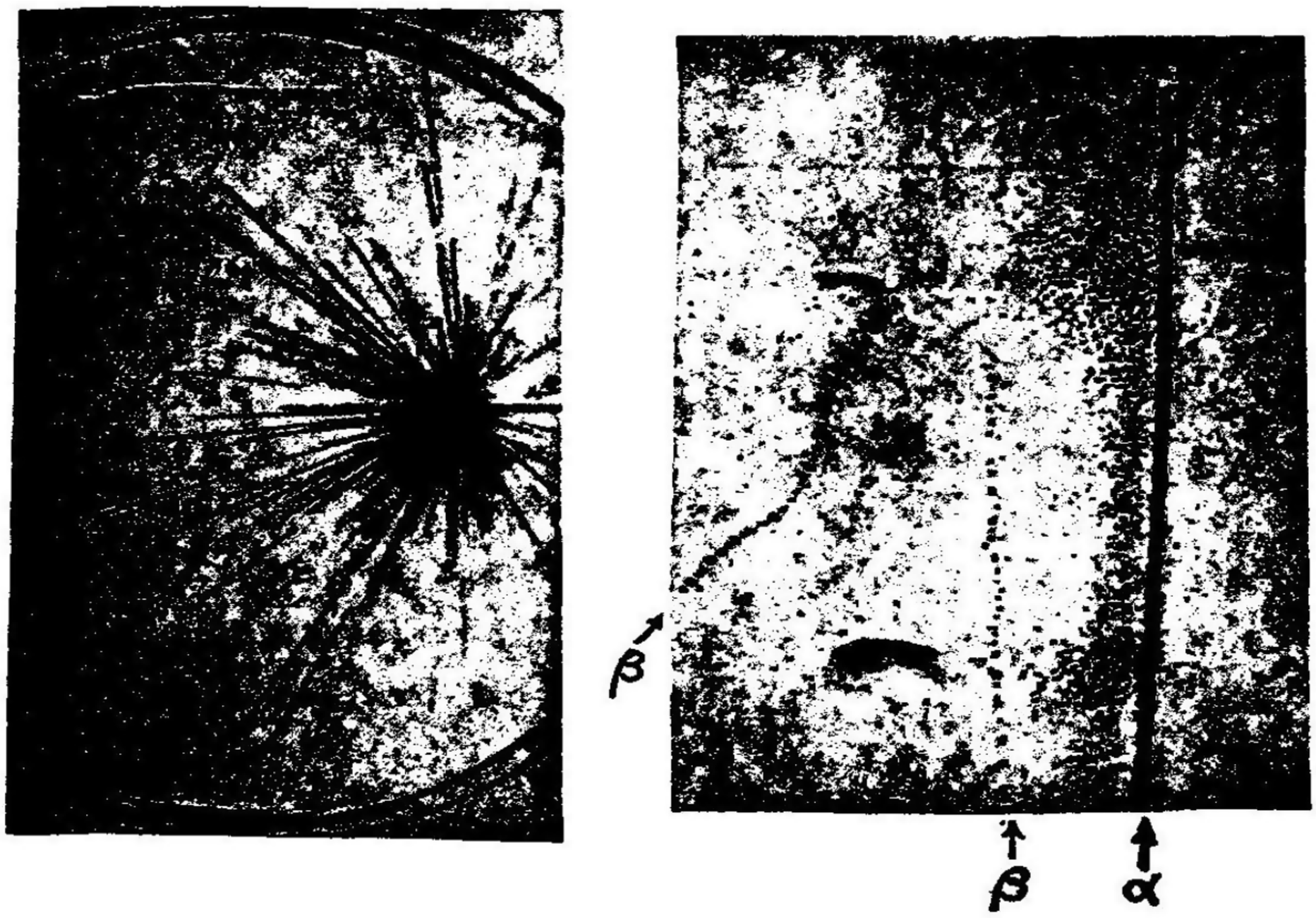


FIG. 156. (a) Alpha ray tracks from radium as seen in a Wilson cloud chamber; (b) One alpha ray and two beta ray tracks. (After C. T. R. Wilson, Proceedings of the Royal Society of London, vol. 87, p. 292, 1912.)

of radiation that when an atom emits an alpha particle its atomic weight is reduced by four, while the emission of a beta particle has essentially no effect on its mass. Similarly the emission of gamma radiation does not change the mass of the atom. Emission of alpha and beta rays should also affect the charge carried by the emitting atom. One of the three known series of radioactive elements, the uranium series, is given in Table 23.

Several points in this table warrant brief mention. The column headed " half life " represents the time required for half of any quantity of that element to disintegrate. A gram of radium today will have become only a half gram 1,690 years from today. By another

1,690 years it will be reduced to half of that, or a quarter gram, and so on. We have already remarked that man has found no means of changing these rates of disintegration, a fact which gives radioactivity its value as a basis for estimating the earth's age. As far as we know now, none of the elements lighter than lead shows the phenomenon naturally, although a few have been given the property by artificial means for short periods of time.

TABLE 23. URANIUM SERIES OF RADIOACTIVE ELEMENTS

Element	*Atomic Weight*	*Ray Emitted*	*Half Life*
Uranium	238	alpha	4.6×10^9 years
Uranium X_1	234	beta	24.5 days
Uranium X_2	234	beta	1.14 min.
Uranium II	234	alpha	3×10^5 years
Ionium	230	alpha	8×10^4 years
Radium	226	alpha	1690 years
Radon	222	alpha	3.82 days
Radium A	218	alpha	3.05 min.
Radium B	214	beta	26.8 min.
Radium C	214	beta	19.7 min.
Radium C′	214	alpha	10^{-6} sec.
Radium D	210	beta	22 years
Radium E	210	beta	5 days
Polonium	210	alpha	140 days
Lead	206	none	infinite

Thus, the existence of radioactivity furnishes additional proof that atoms can be broken down into simpler substances. It also provides still more information about the electron by furnishing another source. Finally, it gives us another building block, the alpha particle. However, as we shall see when we develop a model or picture of the atom, the alpha particle itself can be thought of as composed of simpler parts, and so is a kind of intermediate building block rather than one of the fundamental ones.

Other Atomic Building Blocks

The proton and positron have already been mentioned in connection with static electricity. At about the same time that the positron made its appearance a particle was discovered having essentially the same mass as the proton but neutral in charge. This was

given the name "neutron." It may be thought of as the result to be obtained by closely combining an electron and a proton. Several other particles have entered upon the atomic scene since 1932 but, like the positron, they appear only under special conditions.

When we come to consider models of various atoms somewhat later in this chapter, we shall confine ourselves, for the purposes of this book, to electrons, protons and neutrons as building materials. The following table summarizes their properties.

TABLE 24. PRINCIPAL BUILDING BLOCKS OF MATTER

Particle	*Atomic Weight*	*Charge*
Electron	Essentially 0	−1
Proton	1	+1
Neutron	1	0

THE NUCLEAR MODEL OF THE ATOM

Before we consider the picture of the atom that is usually presented today, there are a few general observations that should be made on the whole problem of atomic models.

Atomic Models in General

The human mind is so constituted that, in general, it prefers the concrete to the abstract. In the field of atomic research, for example, most of us are not satisfied to speak of atoms only in terms of abstract behavior but feel the need of a concrete model that we can make out of beads and wire, place on the desk, and think of as looking like an atom. It should be recognized, however, that an atomic model is a representation of something no one has ever seen and that it must therefore be constructed on the basis of indirect evidence. The builder of the model says essentially, "Here is a group of well-established experimental facts that tell me something of how atoms behave under certain conditions. How can I put together the available building materials in order to get a working combination that could be expected to behave in the same manner?" Two facts should be recognized about this procedure. First, even if we succeed in devising a model that can be shown to fit all the evidence, we still have not proved that it is the only model that could do so. And

second, to be perfect, the model must be in agreement not only with all that is presently known about the atom, but also with everything that will be discovered in the future. Consequently, a model that is entirely satisfactory at one time may prove to be completely unsatisfactory a little later.

As might be expected, a number of models of the atom have been advanced from time to time but have been discarded when new evidence was discovered to contradict them. Unfortunately, the discarding process has often been accompanied by a great deal of bitter controversy. There was a time when chemists and physicists were almost at each other's throats over the question of what the atom looks like. In fact, they behaved very much like the characters in the old fable of the four blind men and the elephant. In that story four blind men were describing an elephant. In examining the elephant one man happened to get hold of its leg and so described it as tall and cylindrical. Another, who had felt its side, pictured it as large and flat like the side of a barn. A third having examined its tail and a fourth its trunk described it in terms to fit their evidence. Each of the four men advanced a model of an elephant that was perfectly satisfactory in the light of his own evidence but was contradicted by the observations of the other three. And so they quarreled. We, of course, think we know how an elephant really looks. At least our model is in accord with the data of all four blind men. Like them, the physicists had an atomic model that agreed with their evidence and the chemists had one that was suggested by their observations, but neither model would conform to the other group's data. Eventually both groups recognized that the only fact their disagreement proved was that neither model was completely satisfactory and that research rather than argument was more likely to produce a model in agreement with all the available evidence.

Basic Assumptions about the Nuclear Model

In describing at least the more common experimental facts known about it, science today thinks of the atom in terms of what may be called a nuclear model. This model, which was first advanced by the Danish physicist, Niels Bohr (1885–), pictures the atom as resembling in appearance a kind of tiny solar system. Correspond-

ing to the sun is a nucleus which, like the sun, contains very nearly all the mass of the system. In the nuclear atom the roles of the planets are taken by electrons. The nucleus is made up of combinations of protons and neutrons. It will be noted that any combination of these two particles will have a net positive charge, since the proton has a charge of +1 and the neutron is neutral. The number of electrons which travel in orbits about the nucleus is just sufficient to cancel this net positive nuclear charge—that is, to make the atom as a whole uncharged.

For example, let us consider a hypothetical atom whose nucleus contains 4 protons and 6 neutrons. Applying the data in Table 24, we see that its atomic weight is 4 + 6, or 10, and that there is a net charge of +4 on the nucleus. There are, accordingly, 4 electrons traveling in orbits about the nucleus. Generalizing, we can say that for any atom the atomic weight is equal to the total number of protons and neutrons in the nucleus, while the number of so-called "planetary" electrons is the same as the number of protons in the nucleus. This method of calculation obviously makes all atomic weights whole numbers. The reason that atomic weights, as usually determined, are not whole numbers will be considered a little later in this chapter.

Although the analogy between the atom and the solar system has been helpful in visualizing the construction of the atom, it is of little importance except perhaps as a method of remembering the general appearance of the nuclear model. In terms of experimental evidence the comparison between the two cannot be carried very far. When this atomic model was first suggested too many people were influenced by a feeling that the idea of a tiny atom being constructed according to the same general plan as the great solar system was too beautiful and inspiring not to be true. The magnificent unity of the illustration impressed them so greatly that they tended to overlook the experimental evidence involved. The place of the solar system in the story of the atom must be regarded solely as analogy and not as evidence.

Examples of Specific Atoms

The composition of six typical atoms, in terms of protons, neutrons, and electrons, is given in Table 25. Referring to it we are now

able to attach physical significance to the term "atomic number" which was mentioned in Chapter 13 in connection with the discussion of the periodic table of elements. In terms of the electrical nature of matter, atomic number can be defined either as the net positive charge on the nucleus, or as the number of planetary electrons in the neutral atom.

TABLE 25. COMPOSITION OF SIX DIFFERENT ATOMS

Element	*Atomic Weight*	*Atomic Number*	*Protons in Nucleus*	*Neutrons in Nucleus*	*Planetary Electrons*
Hydrogen	1	1	1	0	1
Helium	4	2	2	2	2
Lithium	7	3	3	4	3
Neon	20	10	10	10	10
Oxygen	16	8	8	8	8
Iron	56	26	26	30	26

Note that the combination of two protons and two neutrons in the nucleus of a helium atom would give it properties exactly like those attributed to the alpha particle. The identical nature of the two is verified by spectroscopic evidence.

It may also be pointed out that a proton can be defined as the nucleus of a hydrogen atom.

Isotopes

In the preceding table it will be noted that all the atomic weights are given as whole numbers, a situation that follows naturally from considering the nucleus as composed only of protons and neutrons. If the reader will look back at the periodic table as given on page 227 he will see that this is not the case for most elements. Neon, for instance, is given as having an atomic weight of 20.18, not 20 as in Table 25. Now it happens that the usual chemical method of determining atomic weight actually gives an average value for all the atoms in the sample investigated. This, of course, can only be considered to be the weight of each atom if one assumes all of them to be alike. In 1912, using a technique that made it possible to separate atoms of slightly different weights, Sir J. J. Thomson discovered that there are three kinds of neon atoms. In average neon gas about 90 per cent of the atoms have an atomic weight of 20 while the other 10 per cent

are distributed between atoms of weights 21 and 22. When they are mixed together in the usual proportions, the average atomic weight is 20.18. Atoms which thus belong to the same element but differ in atomic weight are called *isotopes* of that element. Since 1912, isotopes have been discovered for almost all of the atoms. Table 26, which is a revised version of a portion of the preceding table, gives the atomic composition of all the known isotopes of the first four elements previously listed.

TABLE 26. LIST OF ISOTOPES OF FOUR ELEMENTS

Element	*Average Atomic Weight*	*Isotope Atomic Weights*	*Atomic Number*	*Nuclear Protons*	*Nuclear Neutrons*	*Planetary Electrons*
Hydrogen	1.008	1	1	1	0	1
		2	1	1	1	1
		3	1	1	2	1
Helium	4	4	2	2	2	2
Lithium	6.94	6	3	3	3	3
		7	3	3	4	3
Neon	20.18	20	10	10	10	10
		21	10	10	11	10
		22	10	10	12	10

It will be recalled that hydrogen of atomic weight 2 was mentioned in Chapter 15 in connection with " heavy water," a compound of H_2O or D_2O in which all the hydrogen is of this type. Hydrogen of atomic weight 3 does not occur naturally but can be made artificially.

The reason that atoms of different atomic weights are still said to be of the same element is that the chemical properties of an atom seem to be determined by its planetary electrons, and all isotopes of a given element have the same atomic number.

ATOMIC ENERGY [1]

We have already stated that under certain conditions energy may be converted into matter and matter into energy. We will now

[1] Much of the material presented in this section was condensed from *Atomic Energy for Military Purposes*, by H. D. Smyth (Princeton University Press and Government Printing Office, 1945). It is recommended that the reader secure access to this reference and read as much of it as possible.

consider briefly the background, mechanism and certain implications of such transformations.

About 1905, Albert Einstein concluded from purely theoretical considerations that mass and energy are related by the equation

$$E = mc^2,$$

which says that if a mass, m, is converted into energy, E, the amount of energy which results equals the mass multiplied by the square of c, the speed of light. Since the speed of light is very great (186,000 miles per second, or 3×10^{10} centimeters per second), this relation indicates that a very small mass is equivalent to a very large amount of energy. One numerical example of this transformation was given in the last paragraph of Chapter 2. Another is that the total conversion into energy of one kilogram (2.2 pounds) of matter would produce 25 billion kilowatt hours of energy, an amount equal to approximately one-sixth the production of the entire United States electrical industry in 1939. In comparison, burning a kilogram of coal in the ordinary way produces only eight or nine kilowatt hours of heat energy.

Although for many years the basis for this mass-energy equivalence remained purely theoretical, the awe-inspiring size of the conversion figure opened an attractive field for speculation. Everyone has seen numerous Sunday-supplement stories on how many trans-oceanic voyages a given ship, usually the current queen of the seas, could make on the energy contained in a teaspoon or so of water. Experimental confirmation of this potential conversion came out of the long series of so-called " atom-smashing " experiments which began as research in pure science long before anyone had contemplated construction of atomic bombs. If atomic nuclei are composed of protons and neutrons, then it should be possible to break up such nuclei into their constituent parts. In other words " atom smashing " means producing artificially somewhat the same kind of transformation that occurs spontaneously when a radioactive element disintegrates.

The technique of disrupting atomic nuclei consists essentially in using protons, neutrons, or alpha particles as projectiles with which to bombard atoms of various kinds. Atom-smashing machines such

as the cyclotron and the Van de Graaff electrostatic generator, described in Chapter 24, are primarily devices for giving these tiny "bullets" enough speed to enable them to disrupt the nuclei with which they collide. The first successful atom-smashing experiment was performed in 1919 by Lord Rutherford when he bombarded nitrogen atoms (*atomic weight, 14*) with alpha particles (*atomic weight, 4*) and produced atoms of oxygen (*atomic weight, 17*) and protons (*atomic weight, 1*). In other words, a few atoms of nitrogen were changed into atoms of one of the isotopes of oxygen, thereby achieving for the first time an artificial transmutation of one element into another.

The Rutherford experiment was the first of a large number of such atomic disruptions which preceded one in 1932 that provided quantitative evidence for the equivalence of mass and energy.

It is a general principle of physics that if a system of any kind is stable, work must be done to break it up. Applied to the present discussion, this means that if an atomic nucleus consists of a stable assembly of neutrons and protons, energy will be required to decompose it into its separate constituents. Or, if we assume mass and energy to be equivalent, we may say that the total mass of a nucleus should be less than the sum of the masses of the individual protons and neutrons which compose it by an amount equivalent to the energy required to hold them together in the nuclear formation. This energy is known as the "binding energy" of the particular nucleus. In the case of a helium nucleus (alpha particle), for example, the sum of the masses of the two protons and two neutrons of which it is composed is 4.033 on the atomic weight scale. The mass of the helium nucleus, however, is 4.003 which differs from the first figure by .03 mass units. This difference is the binding energy of the helium nucleus and represents, of course, an extremely small mass. If, however, we continue to assume that mass and energy are equivalent and apply the Einstein equation we find after multiplying the mass by c^2 that the energy required to break up the nuclei of all the helium atoms in a gram of helium would be about 190,000 kilowatt hours.

In the preceding paragraph, in order to define and illustrate binding energy, we assumed the relation $E = mc^2$ to be correct. Now

we will look briefly at the experiment which first produced actual evidence of this fact. In 1932, experimenters working in Rutherford's laboratory bombarded lithium with high-speed protons (hydrogen nuclei) and found that alpha particles were emitted by the target. The nuclear reaction which occurred can be written,

$$_3Li^7 + {}_1p^1 \rightarrow {}_2\alpha^4 + {}_2\alpha^4,$$

where the superscripts are atomic weights and the subscripts atomic numbers. Careful measurements were made of the mass and of the kinetic (motion) energy both of the target and projectile (lithium and proton, respectively) and of the alpha particles which resulted from the reaction. These showed that a certain amount of mass had apparently disappeared during the reaction while a certain amount of energy seemed to have been produced. It was assumed that if the Einstein equation were correct, multiplying the lost mass by c^2 should give the energy. When this was done the agreement proved to be quite good enough to demonstrate the correctness of the $E = mc^2$ relation. The rate of energy released in this reaction is such that, other factors remaining equal, the combination in this manner of seven grams of lithium with one gram of hydrogen might be expected to produce a half-million kilowatt hours of energy. Again, compare this with the figure given above for heat energy released by burning a kilogram of coal, or with the rate of consumption of electrical energy in your home as indicated on last month's bill.

In view of the results of the 1932 lithium experiment the reader may wonder why the more prosaic sources of energy, such as coal and petroleum, were not quickly superseded by atomic power plants. Briefly, the answer is that prior to the vast research program that led to the atomic bomb all nuclear disruption experiments had been performed with extremely small quantities of material — fractions of a microgram — and in every case the energy required to produce the reaction was much greater than the energy obtained from it. A more complete answer necessitates defining " chain " or " self-propagating " reactions. Ordinary combustion is an example of such a reaction: To burn a pile of shavings we can light a single shaving with a match and the heat released by its combustion is sufficient to ignite neighboring ones which in turn release enough heat to ignite others, and so on.

(The reader will recognize the preceding sentence as a restatement in slightly different form of one of the facts of combustion discussed at the beginning of Chapter 16 on "Fuels.") In releasing energy by the lithium-proton reaction, however, new proton "bullets" have to be provided from an external source to disrupt each new lithium atom. A situation roughly analogous to this would exist if, in burning shavings, each shaving had to be lighted with a separate match; or, to state it another way, if the alpha particles produced by the disruption of a lithium atom had sufficient speed to enable them to disrupt other lithium atoms, producing more alpha particles which could in turn disrupt still more lithium atoms, this reaction would be of the chain variety. In the lithium experiments the alpha particles did not have sufficient energy but it was clear that a reaction of the chain type had to be obtained before any sort of large-scale application could be made of the energy of nuclear disruption.

Eventual achievement of this goal grew out of work on the bombardment of uranium (the most complex of the elements) with neutrons. This reaction was first studied in 1934 and for several years the results proved very difficult to interpret. During the late 1930's, however, a great deal of research on uranium bombardment was carried on and by 1940 the following facts, among others, had been established and were generally known both in this country and abroad:

1. That uranium atoms (atomic weight, 238: atomic number, 92), when bombarded with neutrons, sometimes split into approximately equal fragments which are isotopes of elements in the middle portion of the periodic table. This splitting process is now known as uranium fission.
2. That slow neutrons produce fission in the uranium isotope of atomic weight 235 (U-235) but not in the isotope of atomic weight 238 (U-238), while fast neutrons are less likely to cause fission in U-235 than are slow neutrons.
3. That the energy released during fission of a uranium nucleus is very high.
4. That high-speed neutrons are emitted in the process of fission of the uranium nucleus at an average rate of from one to three neutrons per fission.

Thus, in 1940, everything was ready for a serious attack on the problem of producing atomic energy in quantity, either for controlled atomic power or for an atomic bomb. Mass and energy had been shown to be equivalent and it had been proved that neutrons which caused uranium fission reproduced themselves in the process so that a multiplying chain reaction of the kind described above might be possible. No one yet knew, of course, whether conditions necessary to achieve this goal could be met, but scientists all over the world did have, by this time, a reasonably clear idea of the problems involved.

The principle of operation of an atomic power plant or bomb utilizing uranium fission is implied in the above discussion. One neutron causes a uranium fission that both releases a large amount of energy and causes the emission of more than one additional neutron. Now if these new neutrons are able to produce fission in other uranium atoms, thereby releasing more energy and emitting still more neutrons which in turn do the same, and so on, the desired chain reaction is set up. There are, however, several possible courses which the fission-produced neutrons may follow, only one of which results in producing further fissions. Consequently, whether a chain reaction does or does not occur depends on what percentage of the emitted neutrons follow this one particular course. If the loss of neutrons by the other processes is less than the surplus produced by those following the fission course the chain reaction takes place; otherwise it does not.

In March, 1939, the possible military importance of uranium fission was first called to the attention of the U. S. Government. By that time it could be shown that if all the atoms in a kilogram of U-235 were to undergo fission, the energy released would be equivalent to that set free by the explosion of some 20,000 tons of TNT. It was not then known what percentage of atoms in a uranium bomb might be expected to divide if a chain reaction could be maintained but it was clear that if even one per cent of the theoretically available energy in the bomb of nominal size — say, 200 pounds or less — were to be released explosively, the resulting effect would be of a totally different order of magnitude from that produced by any other known type of bomb. The first government committee on uranium was set up in the fall of 1939. In June, 1940, it was reconstituted under the

National Defense Research Committee (NDRC) and serious investigation was begun on the possibilities of developing an atomic bomb. By the latter part of 1941 the chances of obtaining such bombs for use in World War II seemed great enough to justify an "all-out" effort and a uranium program was set up outside NDRC under the direct administration of the Office of Scientific Research and Development.

Space does not permit discussing here the details of the vast atomic research program which by late summer, 1945, produced the war's most potent and spectacular weapon. It is suggested, however, that the reader look these up in the Smyth report, cited previously.[2] We can, however, mention briefly some of the problems that had to be solved. The fundamental one, as indicated above, was to demonstrate that a self-propagating chain reaction could actually be achieved. This was done first in December, 1942, with normal uranium. The relative probabilities of fission for different isotopes of uranium made U-235 preferable to U-238, which meant that techniques had to be developed for separating the two on a comparatively large scale. Since neutron speeds proved to be critical in the achievement of fission, methods of controlling these speeds had to be worked out. Various procedures were tested for increasing the possibilities that the fission-produced neutrons would produce further fissions. Indications that plutonium (*atomic number, 94; atomic weight, 239*) might be as valuable as pure U-235 for the production of a chain reaction led to extensive studies of means of preparing this element for the purpose. Many technical problems of this sort confronted the scientists engaged in the Manhattan District atomic research program. That they were solved with a considerable degree of success is evidenced by the bombings of Hiroshima and Nagasaki in August, 1945.

Some mention should also be made of the peacetime possibilities of atomic energy. In this connection it is appropriate to quote the following from the Smyth report to which reference has already been made:

"As to the future, one may guess that technical developments will take place along two lines. From the military point of view it

[2] See footnote, p. 503.

is reasonably certain that there will be improvements both in the processes of producing fissionable material and in its use. It is conceivable that totally different methods may be discovered for converting matter into energy since it is to be remembered that the energy released in uranium fission corresponds to the utilization of only about one-tenth of one percent of its mass.

"The possible uses of nuclear energy are not all destructive, however, and the second direction in which technical development can be expected is along the paths of peace. In the fall of 1944 General Groves appointed a committee to look into these possibilities as well as those of military significance. This committee received a multitude of suggestions from men on the various projects, principally along the lines of the use of nuclear energy for power and the use of radioactive by-products for scientific, medical, and industrial purposes. While there was general agreement that a great industry might eventually arise, comparable, perhaps, with the electronics industry, there was disagreement as to how rapidly such an industry would grow; the consensus was that the growth would be slow over a period of many years. At least there is no immediate prospect of running cars with nuclear power or lighting houses with radioactive lamps although there is a good probability that nuclear power for special purposes could be developed within ten years and that plentiful supplies of radioactive materials can have a profound effect on scientific research and perhaps on the treatment of certain diseases in a similar period."

Questions

Discussion

1. Determine the elements for which the following atoms are isotopes and calculate the number of protons, neutrons, and electrons in each one:

a. Atomic weight of 96; atomic number of 42.
b. Atomic weight of 202; atomic number of 80.
c. Atomic weight of 226; atomic number of 88.

2. Look up some accomplishments, other than the ones mentioned in this chapter, of the following people:

a. Robert A. Millikan
b. Madame Curie.
c. Sir J. J. Thomson
d. Heinrich Geissler
e. Sir William Crookes
f. Lord Rutherford

3. Describe the particular experiment with a Crookes tube which indicated that the radiation has a particle rather than a light-like nature.

4. Name all the phenomena mentioned in this chapter which can be called sources of electrons.

5. List all of the following that you could determine, knowing only the number of protons in the nucleus of an atom, and describe how you would go about doing it.

a. Atomic weight.
b. Atomic number.
c. Number of isotopes.
d. Number of neutrons in nucleus.
e. Number of planetary electrons.

6. What basic assumption of John Dalton's atomic theory is definitely contradicted by the phenomenon of radioactivity?

7. Is there any portion of the periodic table for which atomic numbers do not increase as atomic weights increase?

8. Explain how Thomson's experiment proved the cathode particles to be all alike when what he did was to measure an average e/m.

9. Look up in other books something of the work of Madame Curie's daughter, Irene Curie Joliot, and mention some bearing it has on the problems discussed in this chapter.

10. Look up some examples of transmutation other than the two given in this chapter.

11. What properties of the nuclear-type atom would cause it to behave like a tiny magnet and so be a possible answer to the problem raised in the second paragraph of page 440?

12. Define a chain reaction and explain its relation to the practical utilization of atomic energy.

13. Describe the precautions taken during the atomic bomb development program to protect workers against radiation effects. (See footnote reference on page 503.)

14. Look up and read a description of the test at Los Alamos, N. M., of the first atomic bomb; also of the tests conducted at Bikini atoll in the summer of 1946.

Multiple Choice

1. A combination of two alpha particles, four electrons, three beta particles, two protons, and a neutron would as a whole (a) be neutral, (b) have a negative charge of less than 3, (c) have a negative charge greater than 3, (d) have a positive charge of less than 3, (e) have a positive charge greater than 3.

2. Which one of the following is out of place? (a) Cathode rays. (b) Beta rays. (c) Photoelectrons. (d) Alpha rays. (e) Planetary electrons.

3. List all of the following that were true at the time Sir J. J. Thomson began his experiments on cathode rays. (a) Cathode rays had been shown to consist of negatively charged particles. (b) The particles in cathode rays

were known to be all alike. (c) The ratio e/m had been measured. (d) Bohr had advanced the nuclear model of the atom. (e) The effect of pressure on the discharge of electricity through gases had been investigated.

4. An atom of an element that has an atomic weight of 19 and an atomic number of 9 would have (one or more) (a) 10 planetary electrons, (b) a total of 19 protons and neutrons in its nucleus, (c) the same number of electrons in orbits as there are neutrons in the nucleus, (d) a net positive charge of 9 on its nucleus, (e) 10 neutrons in its nucleus.

5. Of the three types of radiation emitted by radioactive substances, (a) all are charged particles, (b) none are charged particles, (c) two are deflected by magnetic and electric fields, (d) all are heavier than hydrogen atoms, (e) two are identical with electrons.

True-False

1. The atomic weight of the combination described in the first multiple choice item would be less than that of a neutral oxygen atom.

2. All members of the uranium radioactive series may be said to be isotopes of each other.

3. The emission of an alpha particle by a radioactive substance should change its atomic number by two units.

4. After an atom of uranium has emitted 5 alpha particles and 2 beta particles the remainder should have an atomic weight of 216.

5. The electrons which Thomson used in his first e/m experiment were cathode rays.

6. At no time in Millikan's experiment did the oil drop acquire or lose more than one electron at a time.

7. The charge of the electron was first calculated by substituting an experimental value for the mass m in the value of e/m as determined by Thomson.

8. An atom whose nucleus contained 6 protons and 6 neutrons could not possibly be an isotope of one that contained 7 protons and 6 neutrons.

9. In the experiment on the bombardment of lithium with protons, as described here, the alpha particles produced during the reaction were able to disrupt other lithium atoms.

10. The principal products of uranium fission are two other isotopes of uranium.

Chapter 27

ELECTRONICS

THE name " electronics " derives from the word electrons, the nature and significance of which have been discussed at some length in Chapters 24, 25 and 26. In the course of these discussions it was pointed out that the electron is the basic unit of negative electricity and that all electric currents consist of electrons in motion. The fields of electricity and electronics might therefore be expected to be closely related. It is true that there is no sharp dividing line between them. The fundamentals of electricity are the fundamentals of electronics, and both are branches of physics. By usage, however, the term electronics has come to be limited more or less to those electrical phenomena which involve the transfer of electrons — that is, the transmission of electricity — through gases or through a vacuum such as occurs in the photoelectric cell or radio tube. For example, a radio circuit is called an electronic circuit while a simple battery-operated doorbell circuit is not so designated.

In this chapter it shall be our purpose first to consider the principles of operation and the outstanding features of the several types of vacuum tubes. Then we will discuss briefly some of the applications of electronics, including two that were of prime importance in World War II. The scope and intent of this book are not such as to warrant detailed consideration of particular electronic circuits; discussion of these may be found in any basic textbook in the field. Several such references are given in the reading list in the back of this book.

VACUUM TUBES

Fundamental to the operation of all vacuum tubes is an electrical phenomenon known as the Edison Effect which consists of the emission of electrons from hot wires. When a filament — for example,

the one found in an ordinary incandescent bulb — is heated by an electric current flowing through it there is a tendency for electrons to be emitted from, or "boil off," the wire. In general, the rate at which this occurs increases as the temperature of the wire increases. This effect is utilized by inserting in the tube one or more other wires, or electrodes, to which the emitted electrons may be attracted under certain conditions. When this occurs a current is flowing through or across a vacuum. Vacuum tubes are named from the number of electric terminals or electrodes they contain. A two-electrode tube is a diode; one containing three electrodes is a triode; while those with four and five electrodes are tetrodes and pentodes respectively. These types will be discussed in order.

The Diode

The principal parts of a two-electrode tube, or diode, are shown in Fig. 157. The component labeled *F* is a filament, the ends of which are connected to a battery; the filament and battery together make up the filament circuit. When the tube is in operation the filament is the negative electrode or cathode. The rectangular metal plate *P* is the other electrode. When the filament is heated by the battery, electrons tend to be emitted as already described. If the plate is attached to the positive side of the battery, as shown here, it carries a positive charge and attracts the emitted electrons to it. Thus there is a second circuit in operation, one in which electrons flow from the filament across the gap to the plate and from the plate through the wires back to the filament again. This is called the plate circuit and includes, it should be noted, a section in which electrons flow across a vacuum without benefit of copper wires, mercury pools, or any other of the common metallic electrical conductors.

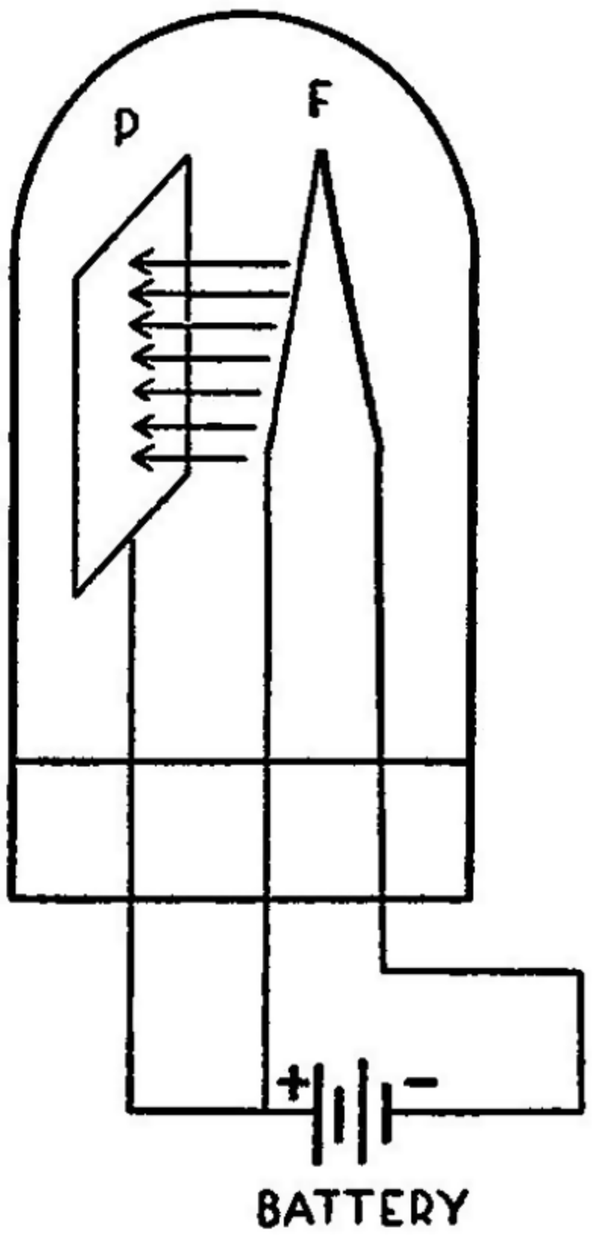


FIG. 157. Schematic diagram of the parts of a two-electrode vacuum tube.

On the other hand, let us suppose the plate is connected to the negative side of the battery and is therefore charged negatively. Since

its charge is now the same as that on the emitted electrons, they are repelled and cannot cross the gap from filament to plate. In other words, if the plate is positive a current flows in the plate circuit; if it is negative no current flows. Thus the tube acts as a kind of electron valve, letting electrons through only when the plate is charged positively. In England the device is called a valve rather than a tube.

The simplest form of cathode is the one shown in the figure in which the filament itself serves as the source of electrons. Among commercial tubes a common variation of this type is an indirectly-heated cathode in which the actual cathode is a cylindrical metal sleeve surrounding the filament or heater wire. The filament is heated by the current it carries and in turn heats the metal sleeve which emits electrons. The cathode may be of nickel coated with a mixture of barium and strontium oxides which greatly increases emission of electrons at a given temperature. In commercial tubes the plate also differs in form from the simple rectangle shown in the figure. It is usually a cylinder or flattened cylinder that surrounds the cathode so that electrons can move directly to it in any lateral direction from the cathode.

Perhaps the most important use of the diode is as a device to rectify alternating currents, that is, to transform alternating currents into direct currents. If an alternating voltage source is connected into the plate circuit of a diode the charge on the plate will alternate between positive and negative. However, current can flow in the plate circuit only during the periods when the plate is positive; during the intervening negative-plate intervals no current flows. Thus, an alternating current is put into the diode but the current that comes out flows always in the same direction and flows only half of the time. Since it does flow only half the time this kind of diode operation is called half-wave rectification. By using at the same time a second diode connected in opposite fashion to the first, the gaps can be filled in with current still flowing in the same direction. In this way so-called " full-wave " rectification is obtained.

The Triode

In the three-electrode tube, or triode, a third element called the *grid* is inserted between the plate and the cathode, as shown in part

(a) of Fig. 158. The grid, which extends the full length of the cathode, usually consists of a mesh of fine wires. Part (b) of Fig. 158 shows the conventional representation of the triode as employed in electrical-circuit diagrams. If the triode is put into operation in exactly the same manner as described above for the diode, and with the grid unconnected, a normal current flows in the plate circuit — that is, electrons move from the filament through the holes in the grid to the plate.

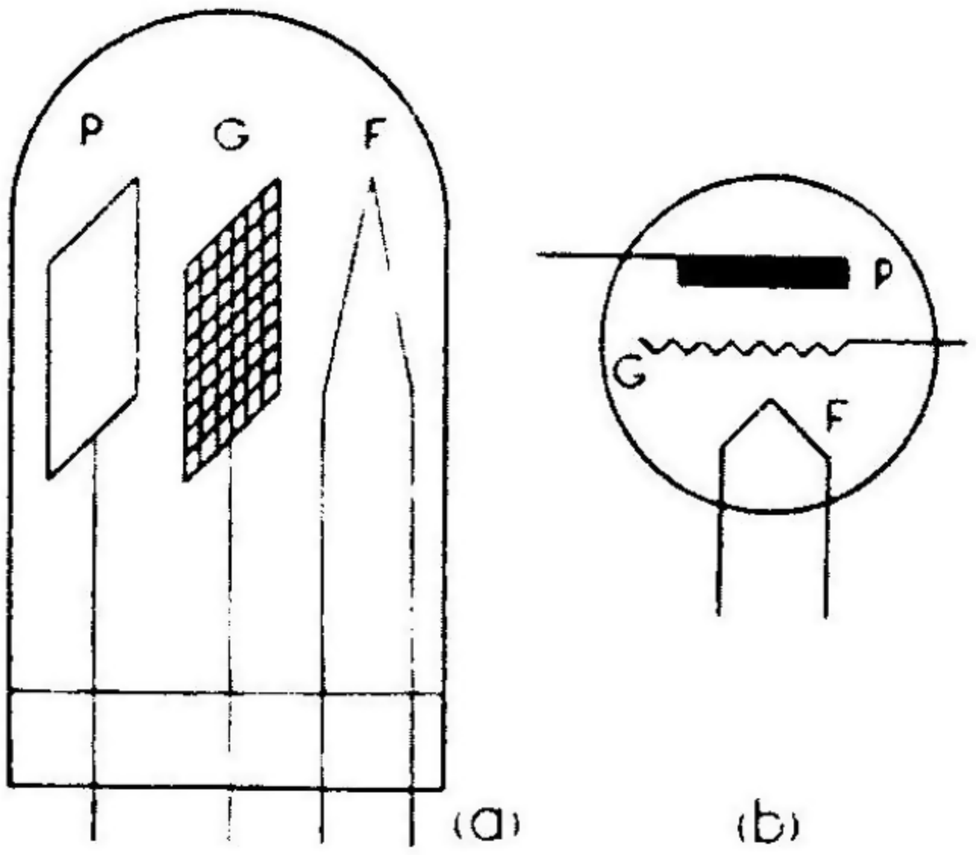


FIG. 158. Schematic (a) and conventional (b) diagrams of the parts in a three-electrode vacuum tube.

The function of the grid is to control this plate current. If a small negative charge is placed on the grid it will repel a few of the electrons and prevent their ever reaching the plate, thereby reducing slightly the plate current; a larger negative charge will cause more electrons to be repelled and

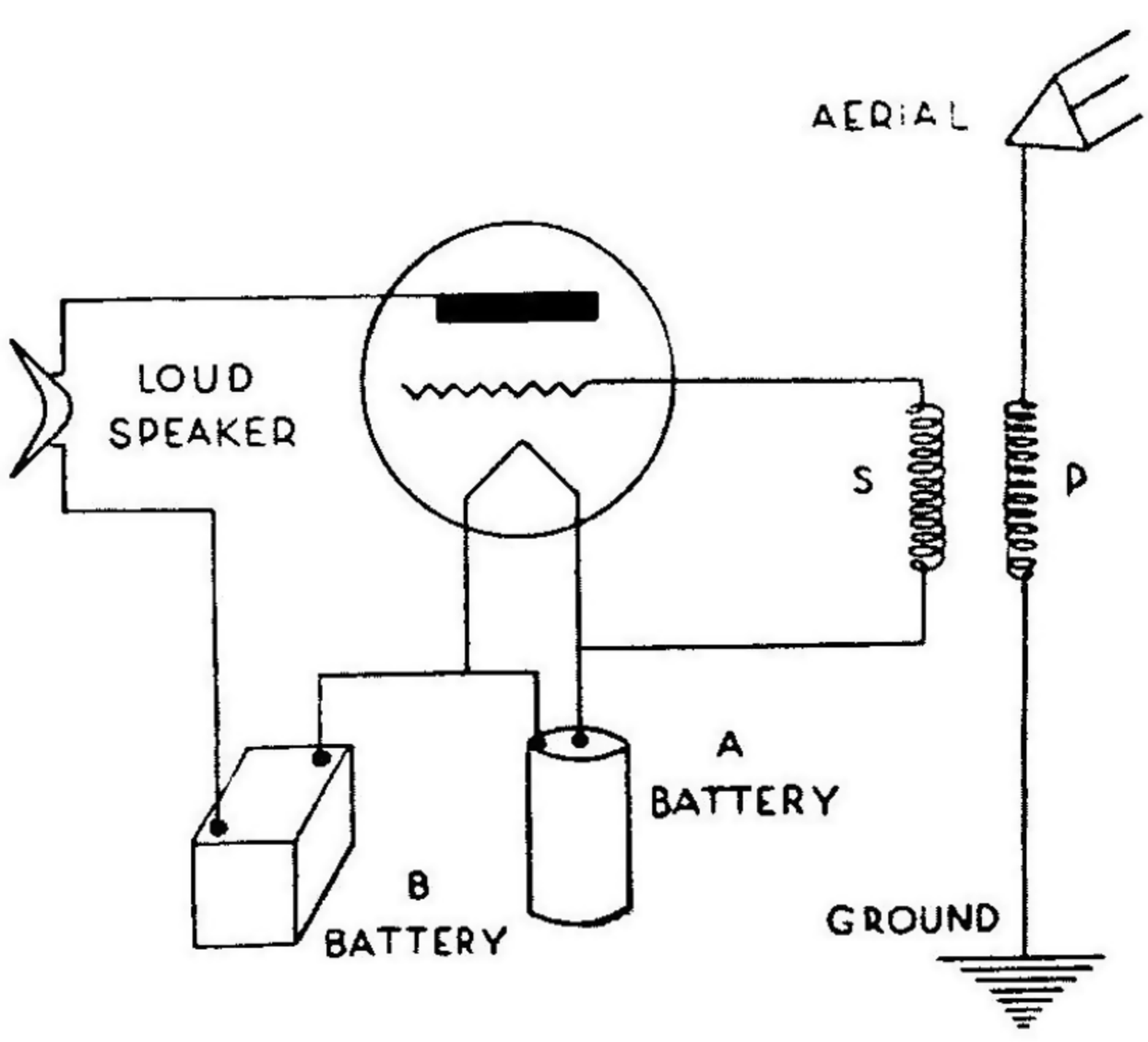


FIG. 159. A simple radio receiving circuit.

reduce the plate current further. In other words, any sequence of variations in the magnitude of the charge on the grid produces a similar sequence of variations in the plate current. The enormous

importance of the triode lies in the fact that *very small variations in the grid charge produce very large variations in the plate current.* To illustrate this point, let us see its application in the simple battery radio receiving set shown in Fig. 159. As the electromagnetic waves sent out by the broadcasting station cut the aerial, they induce a pulsating current in it and in the wire which runs from it through the primary of the transformer to the ground. This varying current in the primary induces variations of similar frequency in the secondary of the transformer. Since the secondary of the transformer is connected to the grid of the vacuum tube the charge on the grid pulsates. This variation of the grid charge produces variations of the same frequency but of greatly increased magnitude in the plate current, as described above. The loudspeaker is placed in the plate circuit. Thus this grid, called the *control grid*, has the effect of greatly amplifying the magnitude of the incoming signals while faithfully reproducing their frequencies. If still greater amplification is desired the plate current of the first tube can be connected through a transformer to the control grid of a second tube.

Tetrodes and Pentodes

In the tetrode the fourth electrode is another wire mesh inserted between the control grid and plate and called the *screen grid.* Its function is to minimize certain so-called " coupling " effects between the cathode, control grid, and plate which limit the range of operation of the triode and may cause it to oscillate or " sing." The new grid is given a positive charge with respect to the cathode so that it attracts and therefore speeds up the electrons traveling from the cathode toward the plate. Since the spaces between the wires in the screen grid are large, almost all of these electrons sift through and reach the plate. Addition of this fourth electrode increases the maximum undistorted amplification that can be obtained at comparatively high frequencies.

Although it solves one problem the screen grid introduces another. As a result of its accelerating action, an increased number of electrons strike the plate with sufficient force to dislodge other electrons. When this happens in the diode or triode all of these evicted electrons find their way back to the plate because it is the only posi-

tively-charged element in the tube. In the tetrode, however, the screen grid is also positively charged and, in addition, is located very close to the plate. Consequently, many of the electrons that are knocked loose from the plate go to the screen grid. Clearly any transfer of electrons in this direction constitutes a decrease in the current through the tube. To solve this difficulty another grid, the fifth electrode, is placed between the screen grid and the plate. Since it is connected to the cathode and so carries the same kind of charge, its effect is to repel back to the plate the slow-moving electrons that were dislodged from it while at the same time not seriously interfering with the regular electron flow from cathode to plate. Since it suppresses the undesired transfer of electrons from plate to screen grid, it is called the *suppressor grid*. The pentode is the tube used in the greatest quantity in radio sets. Still larger undistorted outputs can be obtained with it than with the tetrode.

Tubes containing more than five electrodes are made for various special applications but these will suffice for our discussion here.

Photoelectric Cells

The photoelectric cell or phototube is an electronic device, since its operation involves the transfer of electrons through a vacuum. Since the photoelectric effect as a source of electric current was explained and discussed briefly in Chapter 25, that part of the story need not be repeated here. The photoelectric cell is essentially a two-electrode tube in which the cathode is usually a semicylindrical sheet of light-sensitive metal and the anode is a straight wire placed along the axis of the cathode. A voltage is applied across the two electrodes but no current flows until light falls on the cathode. When this occurs electrons are emitted which travel to the anode because of the applied voltage. In this way the device acts as a switch which can be turned on by shining light on the cathode. Increasing the brightness of the light increases the number of electrons that are emitted but has no effect on their velocities. Velocity of electron emission is affected only by the color, or frequency, of the light, since the velocity is greater the nearer the light is to the ultra-violet end of the spectrum. The group of light-sensitive metals includes rubidium, lithium, potassium, sodium, and caesium. Phototubes are used in setting off burglar

alarms, opening doors, counting, sorting and grading various objects, detecting pin holes in metal, actuating safety devices, in the operation of sound motion pictures, and for other purposes. They can be designed to be sensitive to either visible or invisible light.

Electronic Applications

Several applications of electronics were mentioned in the foregoing discussion of vacuum tubes and photoelectric cells. The electronic devices in use today are so numerous that anything like a complete consideration of them would take us far beyond the space available here. The remainder of this chapter will accordingly be devoted to (a) a general statement of some of the kinds of tasks for which electronic equipment is used and (b) brief accounts of two specific military applications, radar and the proximity fuze. Officials of the War and Navy Departments have ranked these as two of the three most important scientific contributions to winning World War II. The other one is, of course, the atomic bomb.

General Uses

In the field of radio communication the vacuum tube is naturally all-important since it can be said to have made this modern development possible. Certain tubes are employed to generate the frequencies used in broadcasting, while others, as described above, serve to detect the waves and then to amplify them.

So-called ultra-high frequencies are able to generate large amounts of heat in nonconducting materials. Vacuum tubes capable of producing such frequencies have been used with success in the treatment of human ills and for industrial heating. In these therapeutic applications the patient is placed in such a way that heat can be induced within the desired part of the body instead of having to be conducted in as in cases where heat is applied externally. In industry, electronically-induced heat has reduced to seconds the time required to dry painted metallic surfaces and to surface-harden gears, crankshafts, and other machine parts.

One type of tube is used to supply high, direct-current voltages (40,000 to 150,000 volts) where current requirements are low. An example of this application is the removal of dust or smoke particles

from the air by attraction to highly-charged plates as in the " Cottrell Precipitation " described in Chapter 18. Another type of rectifier tube generates, at lower voltages, larger direct currents (20 amperes or less), such as are employed in commercial battery chargers and in the industrial electromagnets used in separating iron and steel scrap from nonmagnetic materials. Still another rectifier tube (gas-filled with mercury-pool cathode) can be used instead of a dynamo to generate comparatively large direct currents. This tube has certain advantages: it possesses no rotating parts, requires no concrete foundation, and is replaceable within a few minutes.

Additional applications of vacuum tubes include their use in voltage regulation where approximately-constant voltage is required for varying currents; in the operation of direct-current motors from alternating-current lines; in the generation of high voltages from low-voltage sources, and for many more specialized functions.

A number of applications of phototubes have already been listed in connection with the discussion of that device.

Radar

In any military operation, whether offensive or defensive, knowledge of the enemy's position is obviously of paramount importance. In the words of the ancient baseball axiom descriptive of the value of the fast ball, " You can't hit them if you can't see them." When applied to any enemy unit — a ship or plane, for example — " seeing " it means determining its bearing (direction) and its range (distance). Radar, a term coined from the words " Radio Detection and Ranging," has enormously increased our ability to locate such enemy units, as well as other objects, because (a) it can " see " farther than the eye even when visibility is good, and (b) the high-frequency radio waves employed in radar are relatively unaffected by darkness, rain, fog, or smoke.

Searchlight Analogy. — As a kind of first approximation to describing the way in which radar works we may draw a rough analogy between the detection of an enemy plane by radar and the location of light-reflecting objects by means of a searchlight. Suppose that a number of objects capable of reflecting light coming from any direction — spherical mirrors, for example — are scattered about a

field and that, at night, you wish to learn their various locations. It will be clear that if you set up a rotating searchlight at your position in the field, each mirror in turn will indicate its location by reflecting a flash of light as the rotating searchlight beam passes over it. In somewhat similar fashion the type of radar used to search for planes or ships utilizes a rotating antenna which sends out radio signals; the process of sweeping the horizon with this radio beam is called *scanning*. As in the case of the mirror, a plane or any other object capable of reflecting radio waves sends back a " flash " of radio energy as the transmitting beam passes over it and so betrays its position. This indicates in a very general way how radar is used to locate objects. Now let us consider the procedure in somewhat greater detail.

Nature of Radar Waves. — Here we must anticipate to some extent material covered in Chapter 31.[1] Radio waves as a whole are a portion of a great spread of waves of different wave lengths known as the electromagnetic spectrum. This spectrum includes, in order of decreasing wave length, radio waves, infra-red rays, visible light, ultra-violet light, X-rays, and gamma rays. If we think of visible light as occupying a position in this spectrum analogous to middle C on the piano keyboard, then gamma rays correspond to tones some 10 or 15 octaves above the highest notes on the piano while the regular broadcast-band of radio waves is analogous to tones more than 20 octaves below the lowest notes on the keyboard. The waves used in radar are longer than infra-red rays but fall toward the short end of the spread of radio waves. All radio waves, as well as the other types of electromagnetic radiation mentioned above, travel at the speed of light, which is 186,000 miles a second. Long and short radio waves differ, however, in the extent to which they are reflected by solid objects. It is characteristic of the ordinary commercial broadcasting band that the waves travel through trees, buildings, and similar obstructions; if this were not so a portable radio with self-contained aerial would not operate indoors. The very short waves used in radar, on the other hand, are reflected by solid objects — a fact which appeared to early experimenters with these waves to be a great disadvantage. The first application of the detection technique which characterizes modern radar came in 1925 when scientists at the

[1] The reader may find it desirable at this point to read the first nine pages of Chapter 31.

Carnegie Institution in Washington used it to measure the distance from the earth to the ionosphere, the uppermost of the atmospheric layers described in Chapter 7. The birth of radar took place when it occurred almost simultaneously to scientists in the United States, Great Britain, France, Germany, and perhaps Japan, that the same technique might be applied to the detection of such objects as ships and airplanes. We may note in passing that the phenomenon of an important scientific discovery being made almost simultaneously by several experimenters working independently in different parts of the world has occurred a number of times. Another example discussed in this book was the technique of separating aluminum from its ore, discovered and patented almost at the same time in France and the United States by two young men neither of whom knew of the other's work. It would seem to demonstrate the futility of hoping to keep any major scientific development a secret for any appreciable length of time.

Operation of Radar. — The searchlight analogy given earlier illustrates in a general way the functioning of radar, but is misleading in one important respect. The radar transmitter emits energy in pulses or bursts [2] which are intense but of extremely short duration rather than continuously, as in the case of the searchlight. Emission of each pulse is followed by a longer period—a few thousandths of a second—during which the transmitter rests while the radar receiver is turned on ready to receive any echoes that may be returned by enemy planes or other reflecting targets. It is this pulse type of operation that makes it possible to determine the range of the target. Suppose, for example, an echo is received one ten-thousandth of a second after emission of the pulse. Since the waves travel with the speed of light, the distance traversed by the energy in going out to the target and back is one ten-thousandth of 186,000 miles or 18.6 miles. If it is 18.6 miles from the radar to the target and back again, then the target must be half that distance, or 9.3 miles away. We have used in this example a target at a range much longer than those frequently encountered in practice. For example, for a target only a thousand yards away the echo reaches the radar receiver six mil-

[2] Actually there is also continuous-wave radar which has special applications not covered in this chapter.

lionths of a second after the pulse was emitted. Clearly the accuracy of the range measurement depends on the accuracy with which extremely small intervals of time can be measured. Today we have learned to determine the position of targets like this within five or ten yards, which means measuring time to an accuracy of about one-thirtieth of a millionth of a second.

Determination of the direction in which the target lies is accomplished in a manner similar to that employed with the searchlight and mirror. The radar antenna must be highly directional; that is, it must send out its pulses in a narrow beam like a searchlight rather than in all directions as in ordinary transmission. In other words, the radar must "narrow-cast" instead of broadcast. Then the direction in which the antenna points when it sends out the pulse that is returned as an echo is the bearing or direction of the target. One type of radar antenna is shown in Fig. 160. This particular radar installation was located near Cologne, Germany, when this picture was taken.

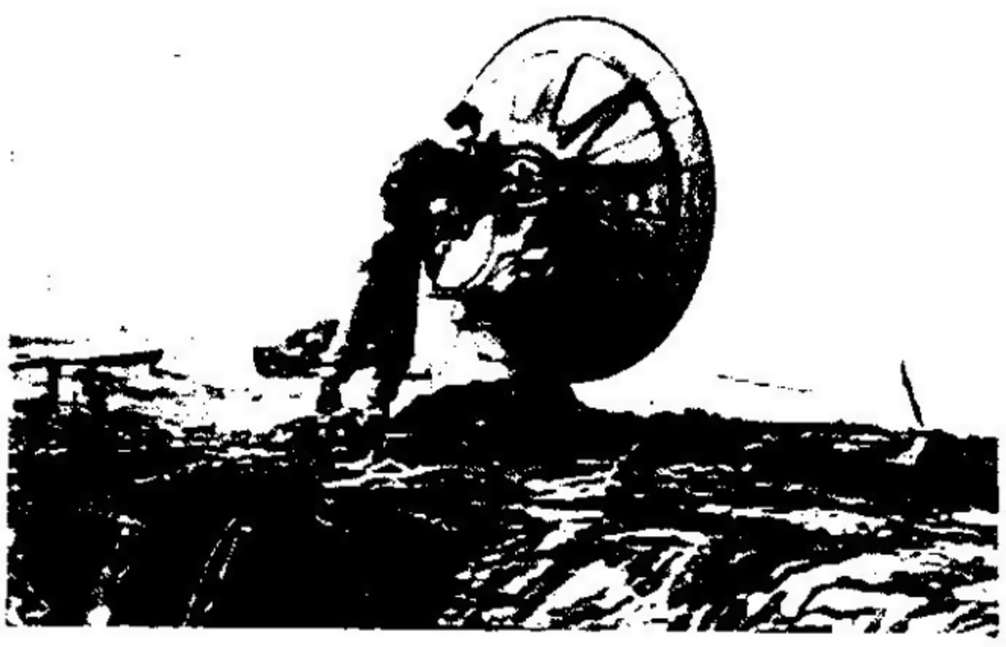


FIG. 160. Dish-type radar antenna being sighted on Cologne cathedral. (Radiation Laboratory, Massachusetts Institute of Technology.)

Thus, by so-called pulse-ranging, both the range and the bearing of the target have been determined. There remains to be described the manner in which this information is conveyed to the operator. A loudspeaker in the radar receiver circuit that would emit a short burst of sound, or "beep," whenever an echo was received might permit reasonably good determination of target bearing but would be almost useless for calculating target range. A much better technique is to make the incoming signal visible and provide a basis for expressing the time interval between emission of the pulse and reception of the echo in terms of distance to the target. This is done by means of a particular type of vacuum tube which is the principal component of an electronic device called an oscilloscope. This tube resembles to some extent the one used by Sir J. J. Thomson in his measurement of the ratio e/m for the electron. (See Fig. 153.) As

in the Thomson experiment the inside of the large end of the tube is coated with a fluorescent material to form a screen. The narrow beam of electrons coming from the cathode is then directed along the tube and produces a spot of light at the point where it strikes the screen. In radar receivers this spot of light is used in several ways to indicate target position. In one type of oscilloscope display, the spot is made to move continuously back and forth across the screen at so rapid a rate that ordinarily the operator sees simply a solid, straight bright line, called the "sweep." Any echo signal that is received is then applied to the tube electrodes in such a way as to deflect the rapidly moving spot at right angles to the sweep to produce a momentary irregularity, or "pip," in the bright line. Since the rate at which the spot is sweeping back and forth is known, the distance of the pip from the end of the sweep is a measure of the distance to the target. The bearing of the target is the direction in which the antenna points when the strongest pip is received.

In another and more spectacular type of oscilloscope display, called the PPI (Plan Position Indicator), the returning echoes draw on the screen what is essentially a map of the region being covered by the radar scanning beam. At the instant each pulse is emitted the electron spot starts from the center of the screen and moves out to the edge along a particular radius which corresponds to the direction in which the antenna is pointing at that instant. If, for example, true bearing is to be depicted on the screen, when the antenna points north the spot travels from the center up to a point corresponding to 12 on a clock face; for an antenna bearing of east the spot moves from the center to 3, and so forth. Thus as the antenna rotates, the electron spot traverses successively a series of radii extending in all directions from the center of the screen. When the PPI is being used to search for planes or ships the intensity of the spot is turned down just at or a little below the point of visibility so that when no echoes are being received the screen face is fairly dark. In this case, the electronic circuits are such that when an echo is received its energy is used to brighten the spot rather than to deflect it. Since the so-called "dark" spot is already barely visible, this echo-brightening is sufficient to make it considerably more so, and the operator sees a bright point of light on the screen at a position which corresponds

to the target's location. For each target sending back an echo, the bearing and range are indicated respectively by the direction and

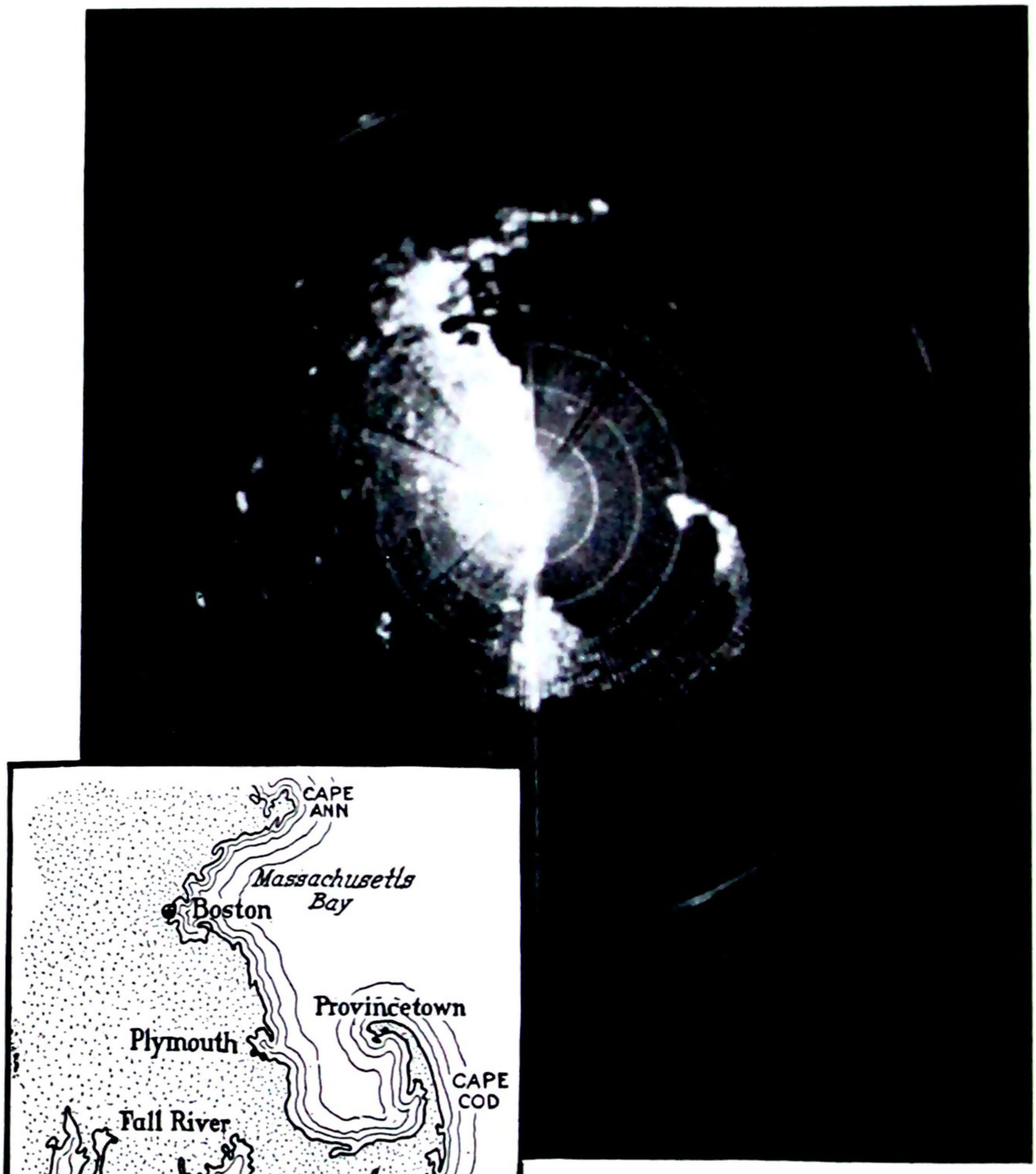


FIG. 161. Radar PPI scope display of Cape Cod region and map for comparison. (Radiation Laboratory, Massachusetts Institute of Technology.)

distance of its particular bright spot from the center of the screen. The persistence of the fluorescence is such that the target spot remains visible longer than the time required for the antenna to make one

complete revolution. Consequently the radar operator has in front of him a continuing map upon which he can see the positions and follow the movements of all targets in the area under investigation. A photograph of the PPI screen in a radar plane flying over Cape Cod is shown in Fig. 161, along with a map of that region for comparison.

The preceding examples concern two of a number of ways of displaying radar-obtained information. If exact range is desired, as on a battleship for use in firing 16-inch guns, the display is such as to permit the operator to read range within a few yards. The radar is so sensitive that the operator can "watch" the shells move across the screen toward the target spot, see the indication of their splashes in the sea and read off the necessary correction. Then on the second salvo he may see the shells land squarely on the target, following which the target spot gradually disappears from the screen. Such incidents happened many times during World War II. In the case of antiaircraft fire the radar antenna may move automatically so as to remain pointed at the enemy plane regardless of its maneuvers and automatically direct the antiaircraft guns accordingly by remote control. Other types of radar make use of other types of display which are designed to accomplish particular purposes.

Applications of Radar. — Since radar has been largely a wartime development, nearly all of its applications to date have been associated with the military field, although some of them have obvious peacetime uses. Among these wartime radar developments are the following:

1. Defense against air attack. This was radar's first active role. It has served both to provide information concerning the approach of enemy planes and actually to direct antiaircraft fire. Some of its most spectacular and important successes occurred during the battle in England against the V-1 robot or "buzz" bombs. Toward the end, the great bulk of these were being shot down by antiaircraft with scores as high on cloudy days, when radar control had to be used, as on sunny days.
2. Submarine location. Airplane-borne radar had a major part in winning the submarine war by facilitating the location of

U-boats at night when they were on the surface charging their batteries. It was limited, however, by the fact that radio waves do not travel any appreciable distance under water, so that submerged submarines could not be detected by radar.

3. Bombing at night and through overcast. Since the passage of radio waves is in no way obstructed by darkness, fog, or clouds, normally invisible targets could be bombed by following the images of the terrain below on the PPI screen. Figure 162 shows a radar screen image of a convoy of ships. One illustration of the importance of this application is the fact that, using radar, the Eighth Air Force dropped 24,000 tons of bombs during December, 1943, while almost no bombing had been possible during the preceding December because of bad weather.

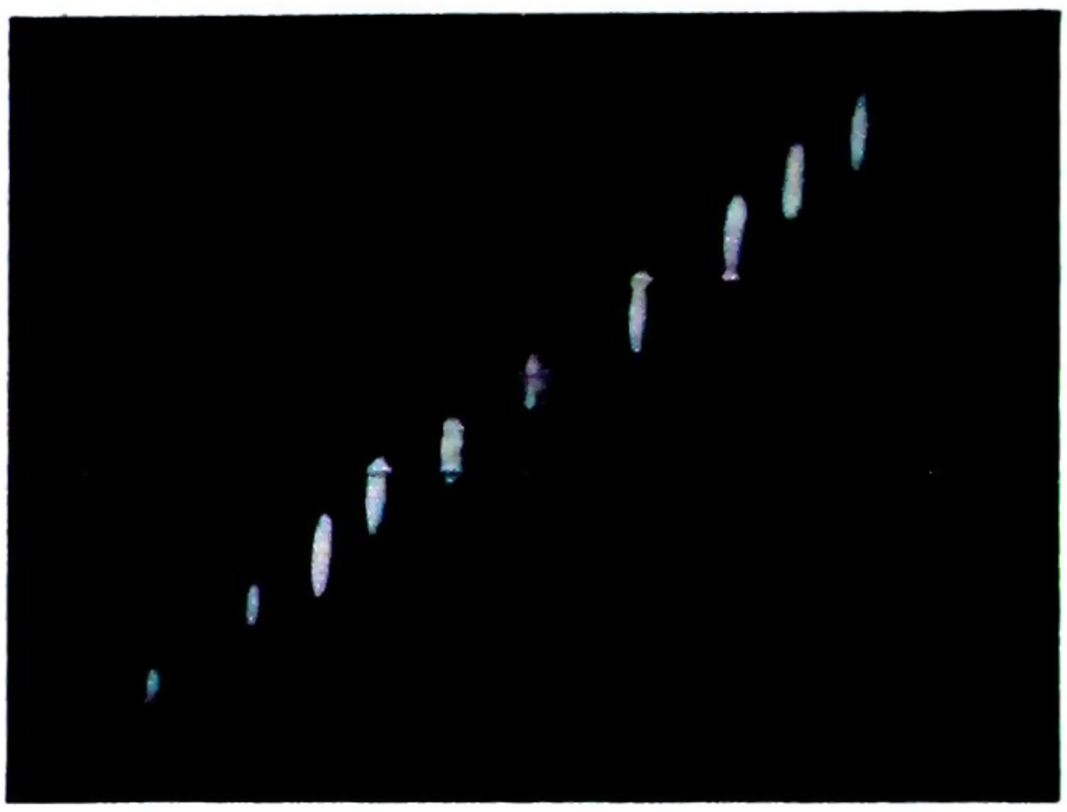


FIG. 162. Radar scope display of ship convoy. (Bendix Radio.)

4. Ground control of approach. A ground-based radar operator, by watching the image of a plane on the PPI and instructing the pilot by radio telephone, can "talk the plane down" to a successful landing on the runway even though the airport is dark or fogbound. A similar technique is employed in direction of fighter aircraft from the ground or a ship.
5. Aid to navigation. One of the most obvious applications of radar is its use on ships or planes to indicate the location of other ships or planes or of natural obstructions such as reefs, icebergs, mountains, and so forth. An example of this use of radar occurred rather early in the war when a group of cruisers, maintaining column formation, were able to steam at 25 knots, with zero visibility, along an intricate path among the islands of a Pacific archipelago, carry out a bombardment mission and return. Back at base they reported that radar information

showed the location of one of the reefs on their maps to be in error by six miles.

6. Radar contact with other celestial bodies. This is being written a few weeks after the U. S. Army Signal Corps announced its first radar contact with the moon. A photograph of the radar scope screen, made during this test, is shown in Fig. 163. The large "pip" on the left indicates the start of the radar impulse toward the moon and the one on the right the return of the echo. The time interval between them was about 2.5 seconds — the time it took the signal to make the trip from the earth to the moon and return at a speed of 186,000 miles a second. The small pips in between the two just mentioned resulted from minor interferences. What practical value such contact may have in the future is yet to be demonstrated; its uses will undoubtedly be fewer and, on the whole, considerably less sensational than those shouted with such enthusiasm from the newspaper and radio housetops the day after the announcement. The chief significance of the experiment at present is the fact that for the first time we know for certain that a high frequency radio wave sent out from the earth can penetrate the electrically-charged ionosphere. Such information is much less exciting than the idea that somehow or other radar contact with the moon might bring closer the day of lunar-bound tourists; it is, however, much more important. A possible future application, after radar has been developed considerably further, is that it might aid in the detailed topographical mapping both of the moon and of certain planets and in determining whether or not life exists on any of the latter group.

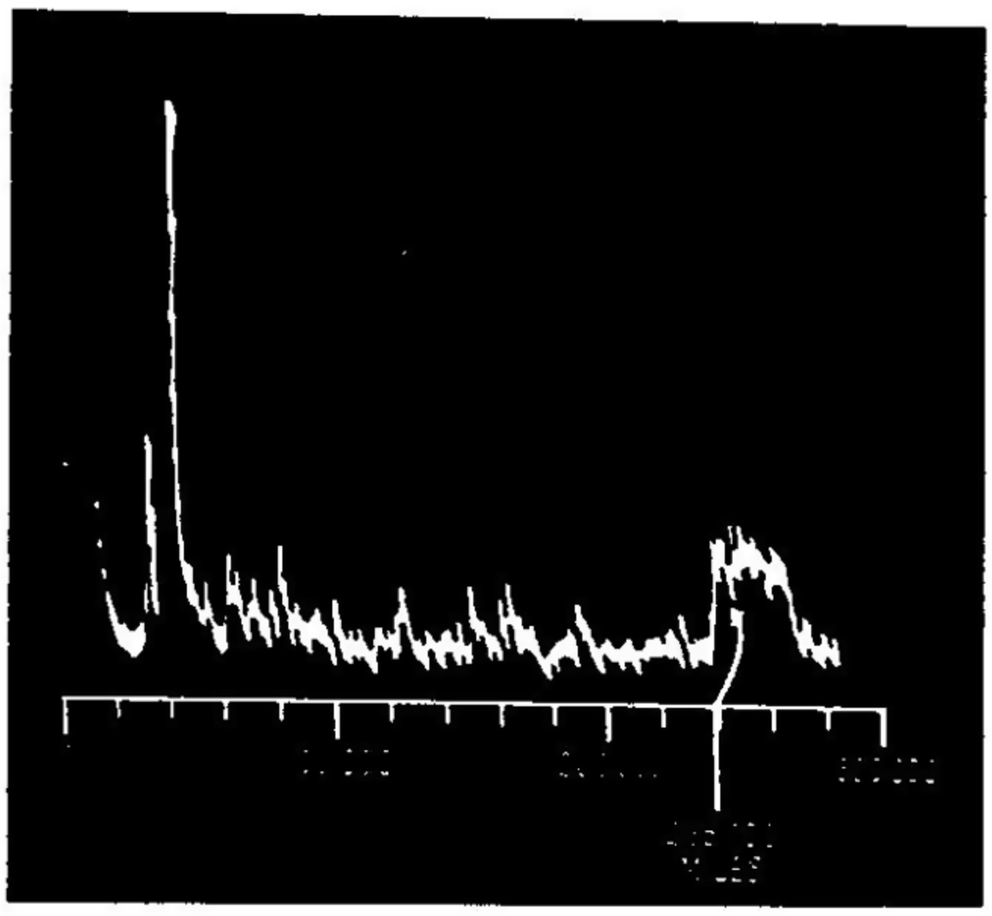


FIG. 163. Scope display of radar contact with moon showing start of transmitted pulse toward moon and reflection of "echo" from it.

These are but a few representative examples of the applications of radar during the war and at present. Many are undoubtedly already familiar to you while its peacetime possibilities will become increasingly clear as time goes on.

The Proximity Fuze

The fuze is that part of a shell which detonates or "sets off" the main explosive charge. The two principal types in use at the beginning of World War II were the time fuze and the contact fuze. The time fuze detonates at a specified time interval after it leaves the gun; consequently, the time of flight from the gun to the target has to be calculated before firing and each fuze "set" for that time. The contact fuze detonates only when it hits something and so is effective only if the accuracy of fire is such that the shell actually strikes the target. It became evident soon after the war in Europe started that the effectiveness of antiaircraft gunfire using either of these fuzes was very low because of the speed, maneuverability, and heights attainable by modern combat airplanes. Attention was accordingly turned to the development of a type of fuze which would be set off by some influence of the target itself if the shell should come within the general vicinity of that target. The idea of such an influence or proximity fuze was not new but had already been proposed independently by various persons both in the United States and in other countries long before the war. There had always seemed to be, however, insurmountable obstacles in the way of making a fuze of this type actually function. The *proximity* or VT (variable time) fuze developed in the United States and first employed against the enemy in 1943 is proof that these obstacles were overcome.

The proximity fuze operates as a *miniature 5-tube sending and receiving radio station* contained within the nose of the shell. The two parts of Fig. 164 show respectively a photograph of one model of the fuze and a cutaway drawing with the parts labeled. When the shell carrying a fuze leaves the gun this tiny broadcasting station begins to send out a radio signal which differs from that described previously for radar in that it is continuous rather than a series of pulses. As the shell approaches its target the signal is reflected back, much as in radar, and picked up by the fuze receiver. The incoming

reflected radio waves interact with the outgoing waves to create a so-called "ripple" pulse which is amplified within the fuze and applied to a particular vacuum tube. In the case of antiaircraft fire,

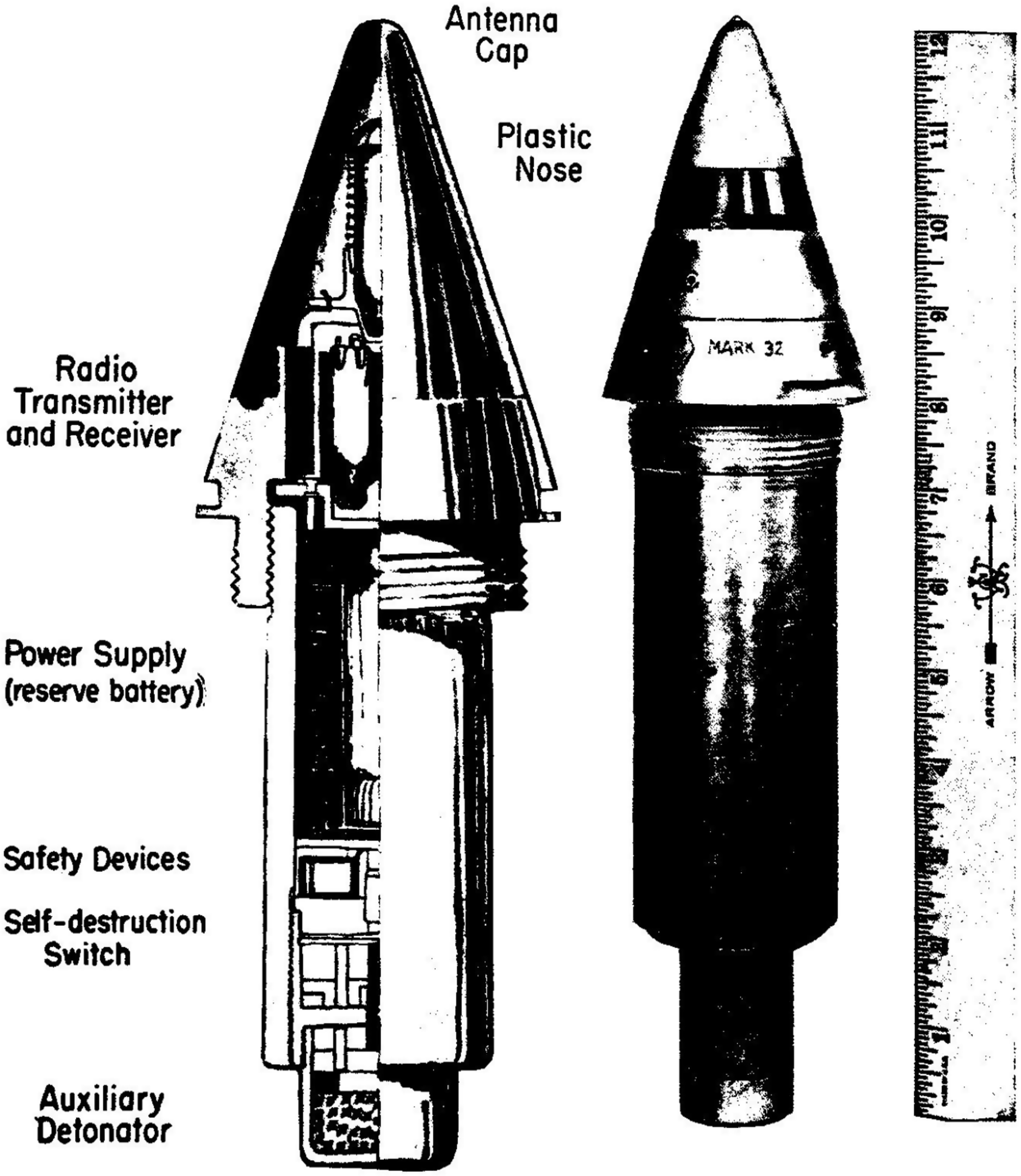


FIG. 164. Cutaway diagram and photograph of the VT proximity fuze. (Applied Physics Laboratory, The Johns Hopkins University.)

when the shell comes within 75 to 100 feet of the target plane the ripple pulse is strong enough to actuate this vacuum tube which serves as a switch to detonate the fuze which in turn sets off the shell.

The proximity fuze contains the following basic components:

1. *Vacuum Tubes.* These vacuum tubes, smaller than a thimble, have to be rugged enough to withstand being shot from a gun,

which means being subjected to a force approximately 20,000 times that of gravity. A photograph of several of the tubes is shown in Fig. 165. Over 130,000,000 of these were manufactured in five years and complete secrecy maintained as to the use to which they were to be put, although at the peak some 10,000 persons were engaged in production and testing. At the height of production one company was turning out some 500,000 tubes a day, which is about 85 per cent of the pre-war peacetime rate of production in this country by all companies.

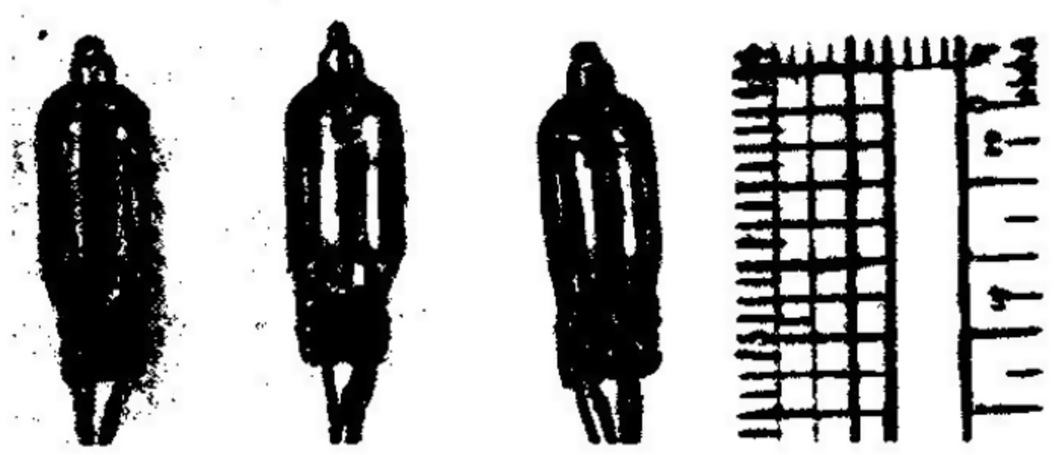


FIG. 165. Miniature radio tubes developed for use in proximity fuze. (Applied Physics Laboratory, The Johns Hopkins University.)

Battery. The battery which serves as the source of electrical energy for the fuze has to be similarly rugged. First successful batteries were of the dry-cell type and were satisfactory except for one disadvantage: they deteriorated with age, particularly in the South Pacific, and this presented a serious storage problem. Later a type of tiny wet battery was developed in which the liquid or electrolyte is held within a glass container in the fuze separate from the battery plates until the shell is fired. The shock of the firing breaks the glass container and permits the liquid to run over the plates when the battery becomes active.

Safety Devices. The principle of operation of the fuze made necessary the development of safety devices that would prevent the unit from functioning until it was a safe distance away from the gun. These were of various forms. Since it was recognized that even with most careful design and construction and the strictest inspection procedures a few faulty safety units — perhaps one per thousand — would get through, each fuze contained two such devices which were entirely independent of each other. Some models of the fuze also carried a "self-destruction" switch which, in case the shell failed to approach close enough to the target for the fuze to function, detonated it after it had passed the target.

The criterion of success of any weapon is, of course, its performance in battle. The proximity fuze was first designed for use in naval 5-inch antiaircraft projectiles because of the urgency of protecting capital ships against air attack. Records of its operation in this field show proximity-fuzed shells to have been three to four times as effective as time-fuzed shells, which is the equivalent, of course, of having three or four times as many guns bearing on the target. When compared with contact-fuzed 40-mm antiaircraft projectiles the margin of advantage for the proximity fuze is even greater. A contact-fuzed shell must strike the target and an approaching airplane presents an actual target area of something like 60 square feet. For the same airplane the area in which a proximity-fuzed shell will be lethal is about 3,000 square feet. In other words for the same enemy plane the gunner using proximity-fuzed shells has a target 50 times as large to shoot at as has the gunner using contact-fuzed shells.

Although the early emphasis on proximity-fuze development was on units for naval antiaircraft shells its use in army projectiles had been contemplated from the beginning. The first battle test of the fuze by the U. S. Army came in December, 1944, during the repulse of Von Rundstedt's counteroffensive in the Ardennes. It was employed with such spectacular success in that major engagement that an official report to the Chief of Ordnance of the War Department characterized it as " the most important innovation in artillery ammunition since the introduction of high-explosive shells."

Questions

Discussion

1. In terms of the definitions at the beginning of this chapter, would the ignition circuit in an automobile be classed as an electronic circuit?

2. What is meant by the Edison Effect?

3. Distinguish between diodes, triodes, tetrodes and pentodes.

4. Sketch a triode and trace both the plate circuit and the filament circuit.

5. What is meant by " rectifying alternating current " and how does a diode accomplish it?

6. What are the functions of each of the following? (a) Triode grid. (b) Screen grid. (c) Suppressor grid.

7. Why is the phototube classed as an electronic device while an ordinary incandescent lamp is not?

8. Look up and discuss several electronic applications not mentioned in this chapter.

9. This chapter was prepared shortly after the first radar contact was made with the moon; discuss any developments in this connection that have occurred since that time.

10. Discuss, both pro and con, the advisability of trying to keep secret an important scientific discovery.

11. If radar contact were to be made with Mars, about how long after the pulse was transmitted should the echo arrive?

12. Describe two different types of screen display used in radar.

13. Find out the difference between true and relative bearing as applied to the display on a radar screen.

14. Distinguish between time, contact and proximity fuzes.

15. List and discuss as many peacetime applications of radar as occur to you.

Multiple Choice

1. The rate at which electrons are emitted by, or "boil off," a hot wire (a) has no relation to the temperature of the wire, (b) increases as the temperature of the wire rises, (c) decreases as the temperature of the wire rises, (d) is zero at all temperatures above 0° C., (e) is constant at all temperatures above 100°.

2. Which one of the following contains no grid? (a) Diode. (b) Triode. (c) Tetrode. (d) Pentode.

3. Under which of the following sets of conditions would the flow of current in the plate circuit of a three-electrode vacuum tube be the greatest? (a) Cold filament, positive plate, positive grid. (b) Hot filament, negative plate, positive grid. (c) Hot filament, positive plate, negative grid. (d) Hot filament, positive plate, positive grid. (e) Cold filament, negative plate, negative grid.

4. Increasing the brightness of the light that falls on a phototube (a) increases the number of electrons emitted per second, (b) increases the speed at which electrons are emitted, (c) affects neither the number nor the speed of electrons emitted, (d) decreases the number, (e) decreases the speed.

5. Radar waves fall within the general classification of (a) cosmic rays, (b) ultraviolet light, (c) radio waves, (d) infra-red rays, (e) X-rays.

True-False

1. In the three-electrode tube, if the plate is made negative, electrons flow from it to the filament instead of in the opposite direction.

2. Current will flow through a diode only when both plate and filament are negative.

3. The function of the screen grid in a tetrode is to screen out the larger electrons and prevent them from reaching the plate.

4. To change the velocity at which electrons are emitted from the cathode of a phototube one must change the frequency of the incident light.

5. Radar waves travel at the speed of light.

6. Radar waves resemble those of the radio broadcast band in that neither is reflected by iron or steel objects.

7. The range of a target can be determined by radar only if a continuous wave is emitted rather than pulses.

8. Radar waves do not travel any appreciable distance through water.

9. Radio waves are sent out by the proximity fuze in pulses rather than continuously.

10. The proximity fuze, developed first for antiaircraft use, also proved effective in artillery shells.

Chapter 28

WAVE MOTION

OUR discussion of waves and wave motion will extend over the next several chapters. In this first chapter we will confine our attention to the more general aspects of the question, that is, relationships and phenomena that apply to all kinds of waves. Then, in the next chapter, we will take up sound waves in particular. Following that we will turn to light and see how it is related to wave motion.

WAVE MOTION AS A METHOD OF ENERGY TRANSMISSION

There are a number of ways in which energy can be transmitted from one place to another. We have already mentioned the use of electrical transmission lines, and pointed out certain advantages that type of energy transfer has over mechanical transmission by belts and gears. Also, we have taken up in some detail two ways in which heat energy travels, namely, conduction and convection. In wave motion we have still another method of accomplishing the transfer of energy.

As an example, consider a chip of wood floating a little distance from the shore on the surface of a quiet pond. Suppose a small boy agitates the water near the shore with his hand. Before long the chip will start to move up and down. Now it possesses kinetic energy, although a few moments ago it was at rest and had none whatever. Let us ask about the source of its energy. Obviously, the energy the chip possesses is part of that which the boy expended. If someone should ask how it gets from the boy to the chip, we would say that it is transported by waves. It is important to note that nothing of a material nature moves out to the piece of wood. The actual molecules of water that are set in motion by the boy's hand clearly do not travel out and strike the chip. All that moves from the boy to the wood is a series of the kind of disturbances that we call waves. One might multiply

such examples almost indefinitely. Suppose one end of a rope is tied to a small tree and you hold the other end in your hand. By moving your end of the rope up and down, the tree can be made to move back and forth. You do work at one end of the rope as a result of which energy is expended at the other end. Again, the particles of rope you hold in your hand do not travel along the rope and bump into the tree. All that travels is a series of waves. Thus, waves possess energy and are capable of transmitting that energy through the medium in which they are moving.

Essentials for the Production of Waves

In every case of energy transmission in which all the observed facts can be accounted for by picturing the phenomenon as a wave motion, two things are found to be present. These are a vibrating source and a medium.

Vibrating Source

In both of the examples mentioned above, the vibrating source was someone's hand. In the one case it vibrated in the water and in the other at the end of a rope. To produce sound waves, perhaps the most familiar of all types of wave motion, something must vibrate. In a tuning fork it is the prongs which vibrate; strings vibrate to produce sound in pianos, violins, and other stringed instruments; in the saxophone it is an air column excited by a reed, in the organ pipe also the source is an air column.

Medium

The necessity of a medium in the first two illustrations that were used is obvious. One would scarcely expect to have water waves without water, or rope waves without a rope. Experiments prove that the media in which sound waves travel must be of a material nature. Suppose an electric bell is set ringing in an airtight jar attached to a vacuum pump. When the jar is full of air the bell can be heard readily. As the air is pumped out the sound becomes fainter and fainter and eventually cannot be heard at all. This means that sound waves will not travel through a vacuum. We may warn the reader now that this matter of medium offers some difficulties when one talks

about light waves. Everyone knows that light will go through a vacuum, so evidently the medium is not a material one. This question will be discussed in more detail in the first chapter on light.

Kinds of Waves

We have already been using one way of classifying waves, namely, into rope waves, water waves, or sound waves. There is also another basis of classification which is of considerable importance. It is one which focuses attention, not on the source of the waves, but on the medium and asks just what happens to a particle of that medium as a wave goes past. From this standpoint waves are grouped into two principal types called *transverse* and *compressional*.

Transverse Waves

Consider the case pictured in Fig. 166. Part (a) indicates the original conditions of the demonstration in which one end of a rope is tied to a tree and the hand holding the other end has just started a single wave moving along the rope toward the tree. Let us fix our attention on the particle of rope labeled x and ask what happens to it as this wave goes past. Three stages in the passage of the wave are shown in the lower part of the diagram. In (b) the particle x has moved vertically upward as the first part of the wave, called the *crest*, has partially passed it. In other words, x is on the crest of the wave. By the time position (c) is reached, x is at the mid-point of the wave and has moved back down to its normal location. part of the wave, called the *trough*, to a point directly below its normal back up to its regular rest position. Every time a wave goes past, moving in a horizontal direction, each particle of the rope partakes of this "up-down-up" motion. Such a wave in which particles of the me-

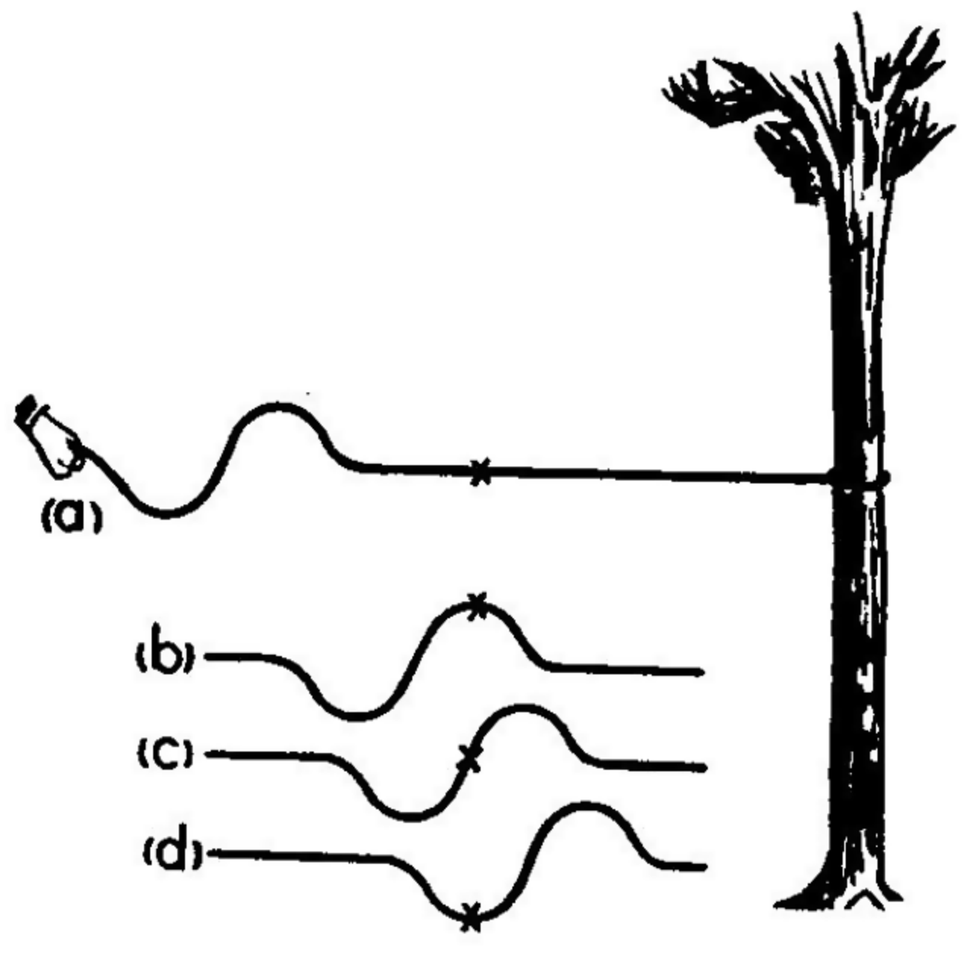


Fig. 166. Diagram illustrating the nature of transverse wave.

In (d), the middle of the second has reached x and drawn it down position. Following this it moves

dium move at right angles to the direction of propagation of the wave is called a *transverse wave*. As we shall see later, there is evidence to indicate that light waves are of this sort.

Compressional Waves

As an aid in visualizing and defining compressional waves, consider Fig. 167. The set-up, shown in (a), consists of a series of ping-pong balls separated by light springs, the whole string of them being suspended between two uprights. Now, as indicated in (b), suppose someone pulls the first two balls together, compressing the spring between them and slightly stretching all the others. Then we ask what will happen when these two balls are suddenly released. It is obvious that the second ball will immediately jump toward the right and then vibrate back and forth eventually coming to rest. The jump to the right compresses the spring between the second and third balls as a result of which the third one is set in motion, also vibrating back and forth. The third one in turn communicates the motion to the fourth and so on until finally a force is exerted on the right-hand support. In part (c) the compression has reached the fourth and fifth balls.

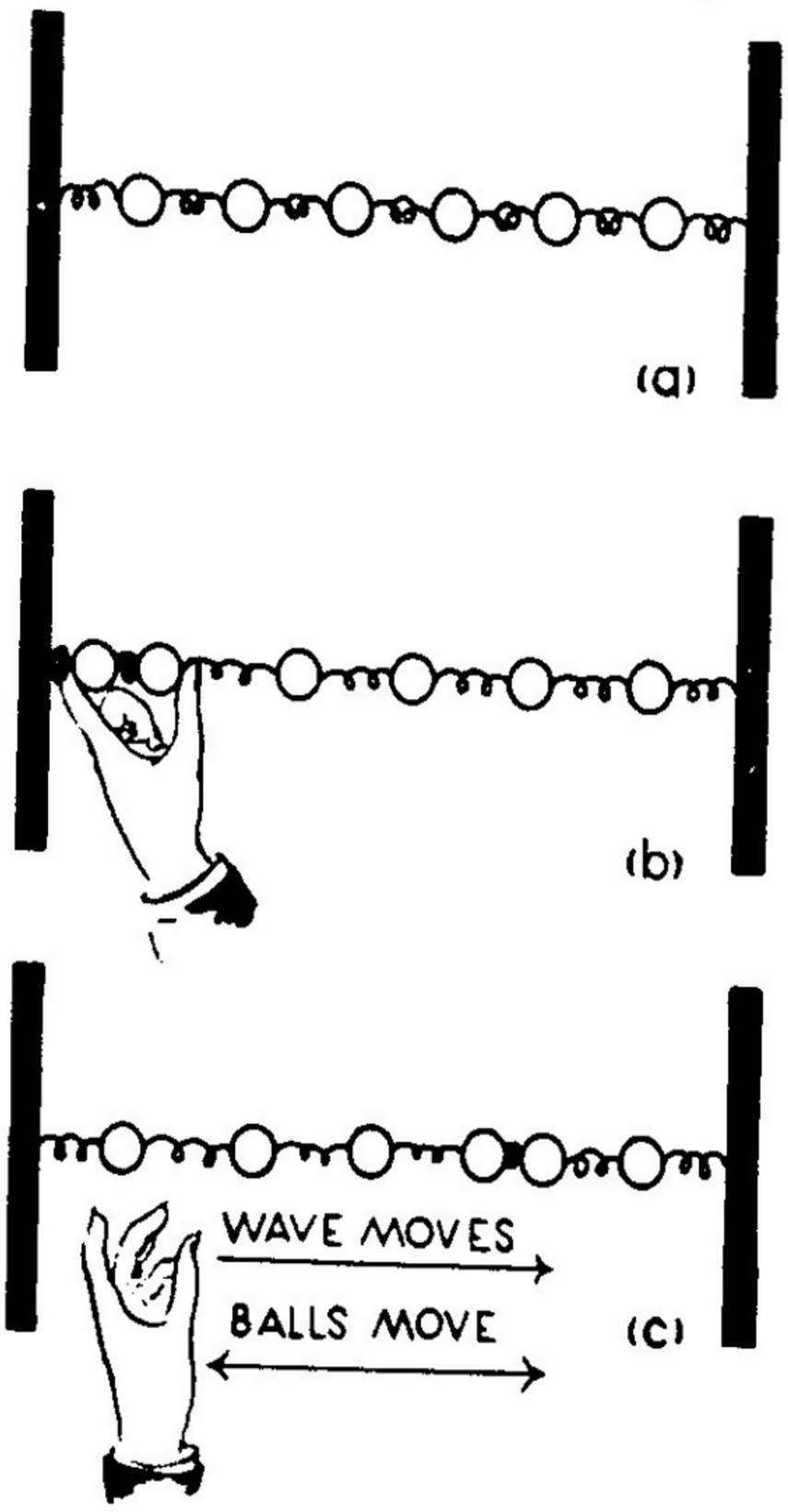


FIG. 167. Diagram illustrating the nature of a compressional or longitudinal wave.

Now note that in this case, as the disturbance moves in a horizontal direction, particles of the medium, the ping-pong balls, also move horizontally. Waves of this sort in which particles of the medium move parallel and anti-parallel to the direction of motion of the wave itself are called *compressional waves* or *longitudinal waves*. A point in the medium that is compressed is called a compression or condensation, while a point where a stretching out occurs is known as a rarefaction or expansion. Sound waves are of this type. Water waves, which were

referred to above, combine characteristics of both kinds in that the particles of water move in small circles as a wave passes, and so have some transverse motion as well as some parallel to the direction in which the wave is moving.

Diagrammatic Representation of Waves

Both transverse and compressional waves are conventionally represented by the type of diagram shown in Fig. 168. If the particular type of wave under consideration actually is a transverse wave, then the two parts are thought of as crest and trough. If the discussion concerns a compressional wave, the same illustration is used with the crest being understood to represent one part of the wave and the trough the other.

Properties of Waves

There are five properties that all waves possess, that are descriptive of various features of the wave itself. Since three of them are closely related to each other we will consider them in one group and the other two by themselves.

Speed, Wave Length, and Frequency

Speed. — All waves possess a finite speed, that is, require time to travel any given distance. This fact is obvious for the rope and water waves previously mentioned. In the case of sound, the facts that the thunder is not heard until some time after the lightning is seen, and that the steam issuing from a boat's whistle is visible a little time before the whistle is heard, prove that sound waves require a finite time to go from one place to another. The magnitude of this speed depends on various characteristics of the particular wave and the medium through which it travels. Wave speeds are measured in the same sort of units as any other speed, namely, a distance unit per a time unit, like feet per second, or centimeters per second.

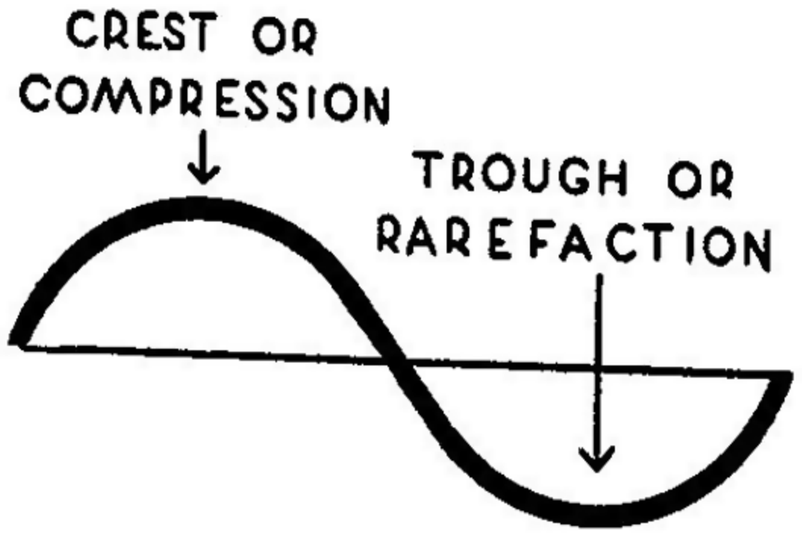


Fig. 168. The conventional wave diagram as used to represent either a transverse or a compressional wave.

Wave Length. — By definition, the wave length in a wave motion is the distance from any point on one wave to the corresponding point on the next one. This is illustrated in Fig. 169 where the wave length is pictured as the distance between crests, between troughs, or the distance from the beginning of one wave to the beginning of the next. As the name indicates it is a length and is measured in distance units like any other length. Variations in wave length in sound and light, the two kinds of waves in which we are particularly interested, affect us physiologically in the following manner. We hear variation in the length of sound waves as variations in pitch, the shorter the wave length the higher the pitch. Changes in wave length in light, we see as changes in color, the violet light waves being the shortest in visible light and the red ones the longest.

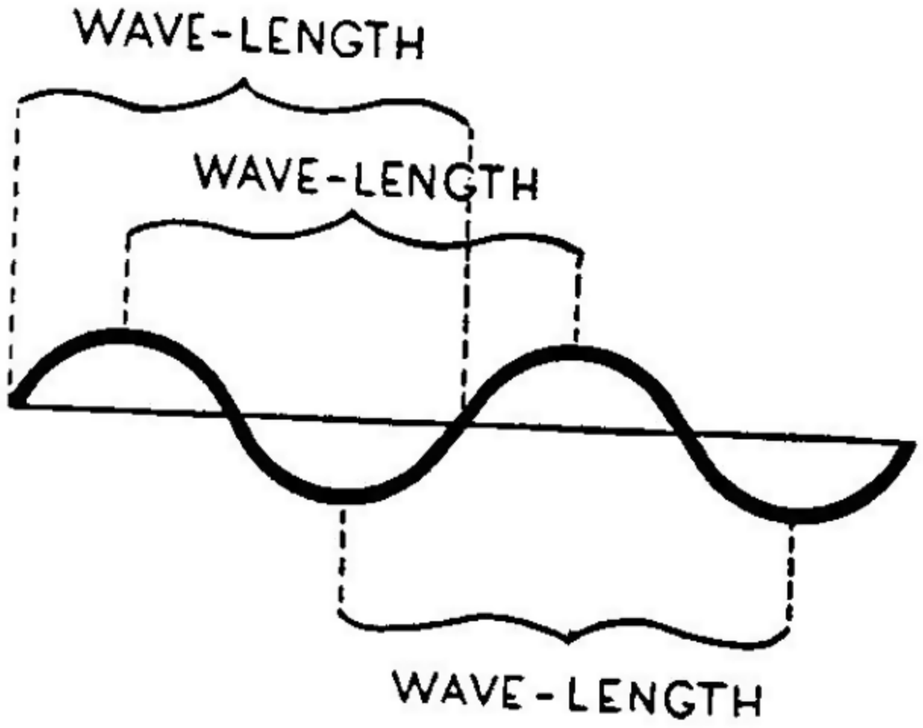


FIG. 169. Diagram illustrating meaning of wave length.

Frequency. — The frequency of any wave motion may be defined as the rate of vibration of the source. It will be obvious from this that it also equals the number of waves sent out by the source per unit time, or the number of waves that pass a given point in the medium per unit time. If the person holding the end of the rope in Fig. 166 makes two complete vibrations with his hand every second, then waves are moving along the rope at that rate and the frequency of the wave motion is two per second. Note that frequency is expressed in units of simply a number per a unit of time. We also hear variations in the frequency of sound waves as variations in pitch, except that in this case as the frequency gets larger the pitch gets higher. Similarly, in the case of light, we see frequency changes as color changes with the violet light having the highest visible frequency and red light the lowest.

Equation Relating Speed, Wave Length, and Frequency. — We had occasion in our earlier discussion of chemical equations to make use of a freight train as an illustration. Let us call your attention again to a string of freight cars, this time as they emerge from a tun-

nel. Suppose you are standing where you can time them and, doing so, discover that cars are coming out of the tunnel at the rate of ten each minute. Then suppose you remember having read somewhere that a freight car is about 40 feet long. A little thought should make it clear that the freight train must be traveling at a speed of 10 times 40, or 400 feet per minute. In other words, if ten 40-foot cars come into sight in one minute, the first one must have moved down the track a distance of 400 feet during that minute. So for the freight trains we can say,

$$Speed = Length\ of\ a\ car \times Number\ of\ cars\ that\ pass\ per\ minute.$$

Now let us translate this illustration into the language of waves. Consider the example already used of the hand that is sending 2 waves per second along a rope. If each wave is, say, 10 inches long, then they must be moving at a speed of 2×10, or 20 inches per second. In other words, for waves,

$$Speed = Wave\ Length \times Frequency.$$

This is an important and fundamental relationship that applies not only to rope waves, but also to water waves, sound waves, light waves, radio waves, and all other waves. We shall have occasion to refer to it in connection with the subsequent discussion of some of these particular types.

Amplitude and Period

Amplitude. — The amplitude of a wave is defined as the maximum displacement from its normal position, experienced by a particle of the medium. In the case of the rope wave in Fig. 166, it would be the maximum distance that point x reaches either above or below its usual position when the rope is straight. The amplitudes of two different waves are shown in (a) of Fig. 170. Being a distance, amplitude is measured in ordinary length units. Fig. 170 also illustrates diagrammatically the difference between amplitude and wave length. The two waves shown in (a) have the same wave length but different amplitudes, while those shown in (b) have different wave lengths and the same amplitude. In sound, variations in amplitude register on our hearing as variations in loudness, with greater amplitudes producing louder sounds. We see differences in amplitude in light waves as

differences in brightness, the greater the amplitude the brighter the light. So if the waves pictured in (a) of the figure were sound waves, the solid one would be the louder, if light waves it would be the brighter.

Period. — The period of a wave motion may be defined either as the time required for one wave to pass a given point or as the time required for the source to complete one vibration. It is related in a simple fashion to frequency. It will be recalled that the frequency was given as equal to the number of waves that pass a point per unit time. Let us go back to the freight train for a moment. Ten cars emerged from the tunnel each minute. That is, the frequency was 10 cars per minute. Obviously, then, it required a tenth of a minute for one car to emerge, or to pass a given point. Thus, the period was 1/10 of a minute, which is equal to 1 divided by the frequency. Or, to take another example, if a certain source of sound has a frequency of 256 vibrations per second, then 256 waves go past a given point each second and it requires 1/256 of a second for one to go past. Generalizing, we can say that for any wave motion,

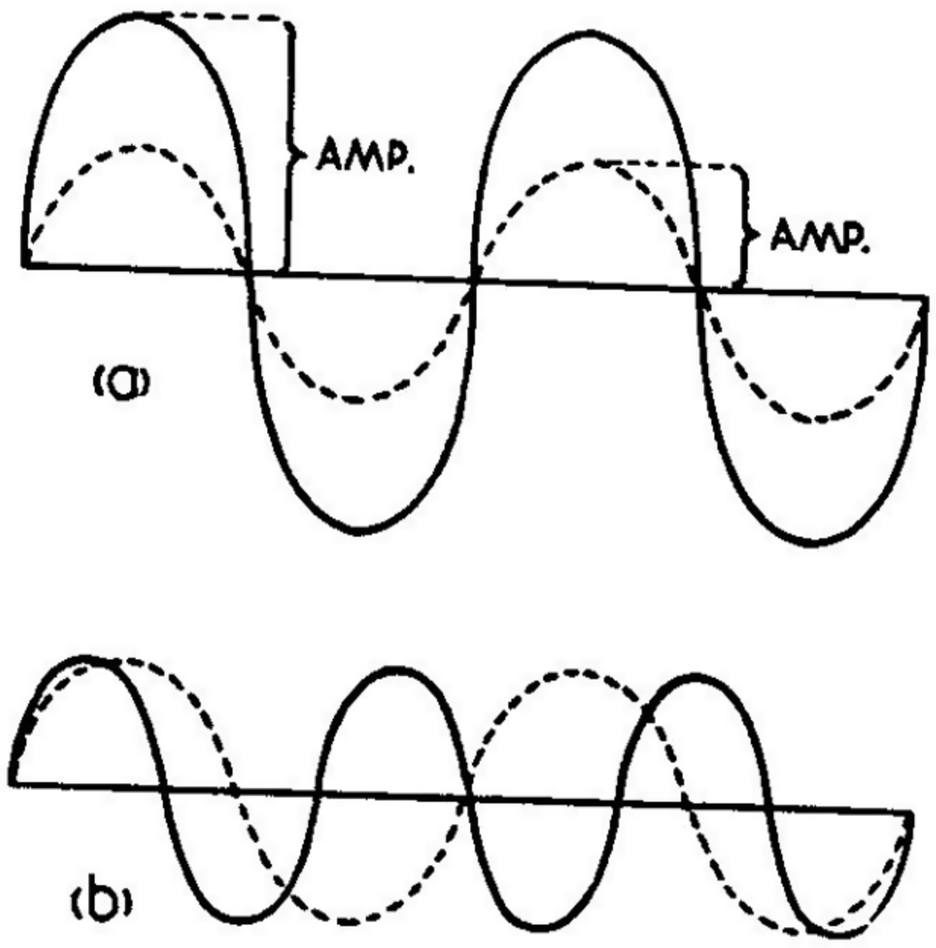


FIG. 170. Diagrams illustrating difference between amplitude and wave-length. (a) Two waves with the same wave-length but different amplitudes. (b) Two waves with the same amplitude but different wave-lengths.

$$Period = \frac{1}{Frequency}$$

Now that we have defined and discussed briefly various general features of waves, we will proceed in the next chapter to a consideration of sound waves in particular.

QUESTIONS

Discussion

1. Discuss the difference between transverse and compressional waves.
2. How might one go about proving that in water waves, the actual water itself is not carried along with the wave?

3. Trace back to the sun the energy acquired by the floating chip used as an illustration in this chapter.

4. In what directions might each man move as a transverse wave passes through a single row of soldiers lined up in an east and west direction?

5. Calculate the period of a rope wave which has a wave length of 4 feet and travels at a speed of 6 feet per second. Use decimal fractions.

6. If the two waves pictured in (*b*) of Fig. 170 represent visible light waves, which one is nearer the red end of the spectrum?

7. If the waves referred to in Ques. 6 are sound waves, which one represents the lower pitch? the smaller frequency?

8. Answer Ques. 6 and 7 for (*a*) of Fig. 170.

Multiple Choice

1. The frequency of a wave equals the (a) distance from one crest to the next, (b) time of one complete vibration, (c) maximum displacement of a point in the medium, (d) wave length multiplied by the period, (e) velocity divided by the wave length.

2. Two sounds that differ in loudness must necessarily have different (a) velocities, (b) wave lengths, (c) amplitudes, (d) periods, (e) frequencies.

3. Suppose a number of men are lined up in single file in a north and south direction and a compressional wave passes along the line; in which of the following directions (one or more) might a man move first when the wave reaches him? (a) north, (b) south, (c) east, (d) west, (e) up.

4. Which one of the following could be the result obtained by multiplying a wave length by a frequency? (a) 1,100 feet per second, (b) .1 second, (c) 2 seconds per vibration, (d) 256 vibrations per second, (e) 4 kilometers.

5. One might determine the period of a wave motion by (a) dividing the amplitude by the frequency, (b) squaring the wave length, (c) multiplying the amplitude by the wave length, (d) dividing the wave length by the velocity, (e) multiplying the frequency by one.

True–False

1. Water waves are an example of pure compressional waves.

2. A wave that had a period of half a minute would have a frequency of 30 vibrations per second.

3. A vibrating source that produces blue light could not possibly have the same amplitude as one that produces red light.

4. One of the two essentials given for the production of waves was a vibrating source.

5. The chief difference between transverse and compressional waves is that actual matter travels through the medium in the latter case but not in the former.

6. In terms of wave diagrams, a trough of a transverse wave is thought of as corresponding to a rarefaction of a compressional wave.

7. Sound is the only type of wave motion that does not seem to require a material medium for its transmission.

8. Light waves travel with an infinite velocity.

9. The property of a sound wave that determines its loudness is the same as that of a light wave that determines its brightness.

10. A wave that had a velocity of 1,100 feet per second and a frequency of 220 vibrations per second would have a length of 5 feet.

Chapter 29

SOUND

IN the first part of this chapter we shall apply to sound waves in a more detailed manner the topics discussed generally in the preceding chapter. Then we shall take up a group of sound phenomena whose manifestations are more or less familiar to most people. Finally, in the latter part of the chapter, the physical basis of that particular division of sound that we call music will be considered.

SOUND AS A WAVE MOTION

Types of Waves

It was stated in Chapter 28 that sound waves are of the compressional type. In order to get a visual image of the actual anatomy of a sound wave as it travels through air, let us consider Fig. 171. It shows

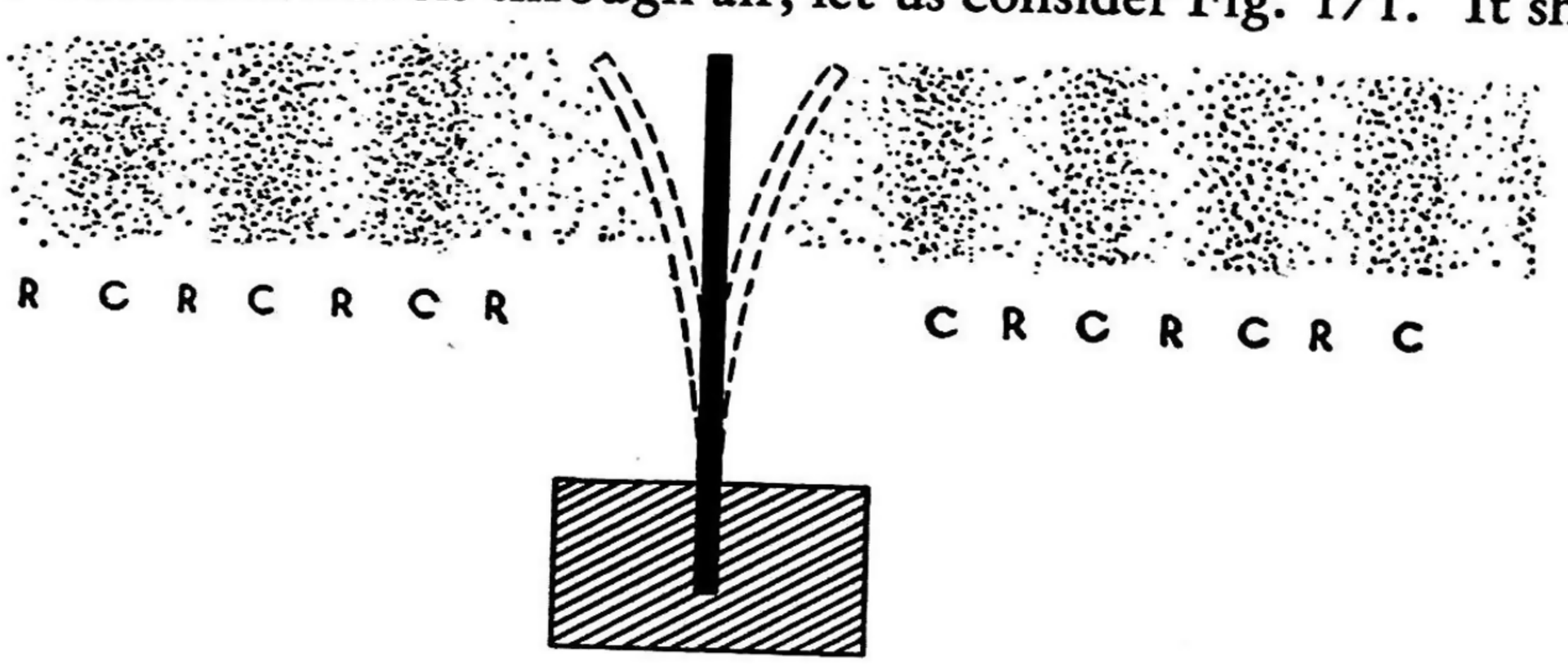


FIG. 171. Diagram illustrating nature of sound waves in air as produced by a vibrating rod. C indicates compression; R indicates rarefaction.

a steel rod or wire, one end of which is tightly held in a block, while the other end is vibrating back and forth. Such an oscillating rod will, of course, emit a sound. Let us trace this rod through one or two complete vibrations. As it moves toward the right it pushes the air molecules in front of it somewhat closer together than normal, creating a compressed region or *condensation* in the air. Then, as the rod goes

back toward the left a partial vacuum is created behind it, which means simply that the molecules of air are pulled farther apart than normal. Such a region is called a *rarefaction.* This rarefaction will be followed by another condensation as the rod moves to the right again and it, in turn, by another rarefaction, and so on. It is obvious that there will be a similar series of condensations and rarefactions on the other side of the rod. When these alternating compressed and rarefied parts of the atmosphere strike our ear drums, they cause them to vibrate at the same rate as the iron rod, and we say we hear a sound. Thus, it is important to note that a sound wave in air consists simply of a particular sort of abnormal distribution of air particles, a distribution in which compressed regions and regions of partial vacuum alternate with each other. When a normal distribution of air molecules exists in the air, no sound waves are traveling through it. Fig. 172 shows how sound waves originating from a point source would appear if one could see individual molecules in the air. Actually, the parts of the waves move out in spheres rather than circles. Before going on to the next section, the reader should compare the picture of a sound wave as described here with the illustration (Fig. 167) used in the preceding chapter when compressional waves were being defined.

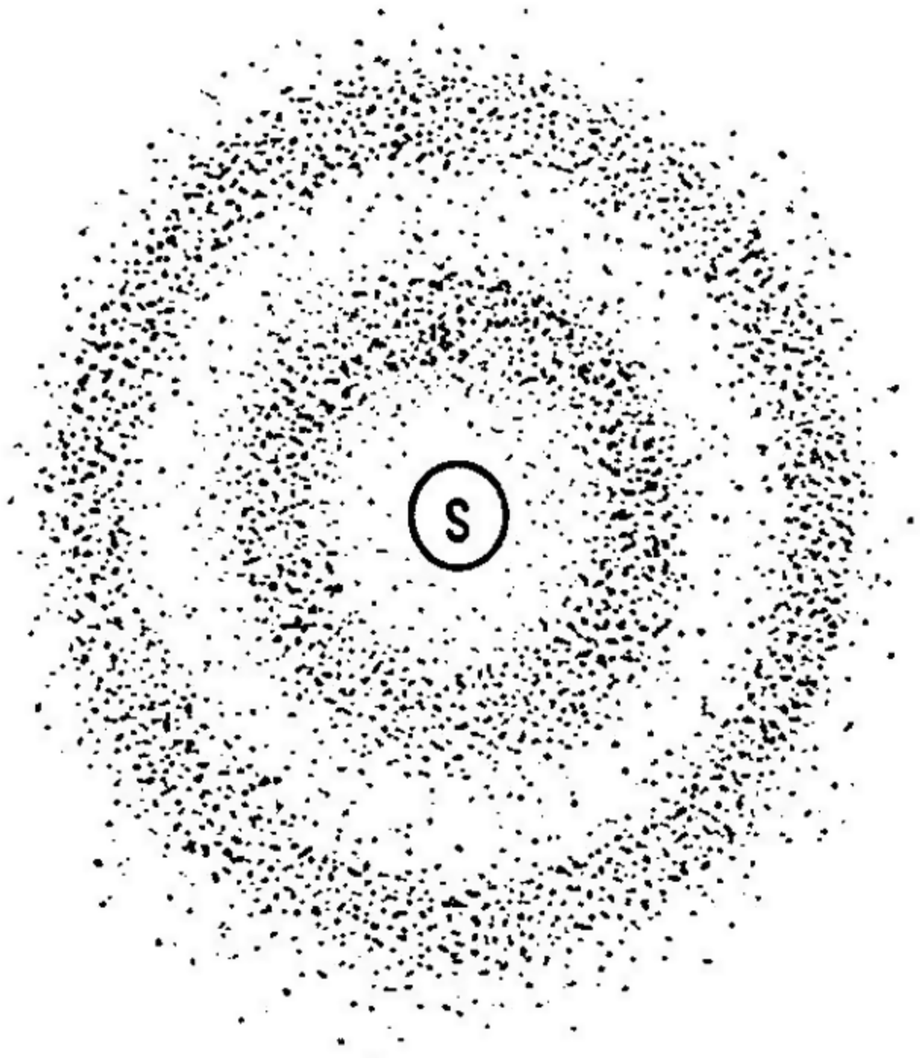


FIG. 172. Diagrammatic representation of a sound wave in air produced by a point source.

Essentials of Sound Wave Production

It will be recalled that sound waves were used as examples in the discussion of the essentials for the production of waves in general. Mention was made of several kinds of vibrating sources as they function in producing speech and in the operation of various musical instruments. That the vibrating source does not have to be anything particularly elaborate or complicated in order to produce a sound may be demonstrated by the simple devices pictured in Fig. 173. The one

on the left consists of a group of toothed wheels. A card held against any one of them will vibrate as many times per revolution as there are teeth on the wheel, producing a definite tone. The greater the number of teeth, the greater is the rate of vibration of the card and the higher the tone. Demonstration pieces of this sort are usually designed to produce the tones *do*, *mi*, *sol*, and upper *do*. The device on the right consists of a disk having several sets of evenly spaced holes arranged in circles around it. A jet of compressed air blows through each hole of a set in turn as the disk rotates. The intermittent pulses of air produced in this fashion cause a sound to be emitted. Again, the more pulses per revolution the greater is the frequency and the higher the pitch of the tone.

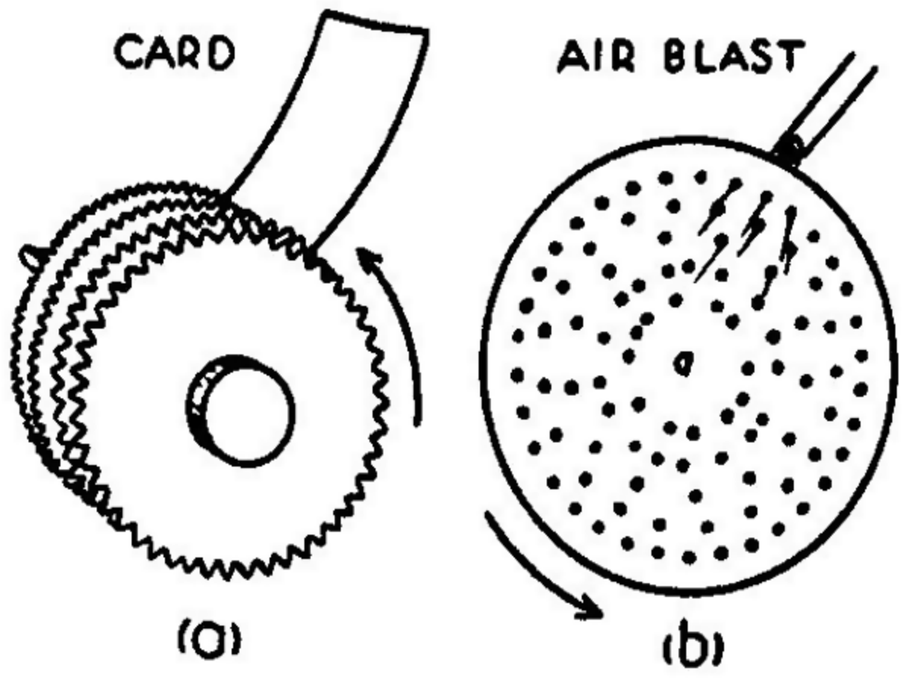


FIG. 173. Examples of devices with which the relationship between frequency and pitch may be demonstrated. (a) a toothed wheel. (b) a disk with evenly spaced holes arranged in concentric circles.

Properties of Sound Waves

Speed. — Table 27 lists the approximate speed of sound in a number of media. For purposes of any calculations we will make in this book, the speed of sound in air at ordinary temperatures may be taken as about 1,100 feet per second.

TABLE 27. SPEED OF SOUND IN VARIOUS MEDIA

Substance	*Speed in Feet per Second*
Air (32° F.)	1,090
Glass	16,000
Hydrogen	4,160
Ice	10,500
Iron (cast)	15,450
Oxygen	1,040
Rubber	100
Steel	16,800
Wood, Ash	
Parallel to grain	15,310
Across grain	4,570

Two important properties of the medium which affect the speed with which sound goes through it are elasticity and density. The mathematical equation relating speed to these properties need not concern us in this book. Qualitatively, the relationship is simply that, other things being equal, speed increases with increased elasticity but decreases with increased density. However, the relations are not direct and inverse proportions respectively. The changes in elasticity and density of gases with rise in temperature combine in such a manner that speed of sound increases as the temperature of the gaseous medium gets higher regardless of whether the gas is free or confined in a vessel. In air the increase amounts to about 1.1 feet per second per degree Fahrenheit rise in temperature. It should also be mentioned that the speed of sound in the atmosphere is approximately independent of loudness and pitch. That this is fortunate is evident when one stops to think how an orchestra concert would sound to someone seated in the back of the balcony if, for instance, the loud tones traveled much faster than the faint ones and if the high notes of the piccolo had a speed several times as great as the lower pitched ones emitted by the tuba.

Frequency. — It has already been pointed out that we hear changes in frequency of sound waves as variations in pitch, with the pitch rising as the frequency increases. There are both upper and lower limits to the frequencies that are audible to the human ear. Experiments show that the average person hears frequencies below 15 or 20 per second as individual impulses rather than as a continuous sound and that 15,000 to 20,000 vibrations per second are the highest that he can hear. The actual values of these limits vary appreciably with individuals and even with the same person at different ages, or with different conditions of health.

Animals other than man are able to hear frequencies that are quite inaudible to human beings. The fact that dogs are able to hear frequencies higher than 20,000 makes possible the type of dog whistle that is unheard by human ears but is quite satisfactory as an instrument for calling the dog.

Wave Length. — Wave length has already been defined as the distance between corresponding points on adjacent waves. Making use of the equation introduced in the preceding chapter, we can calculate

the actual length of the waves which make up that best known of all musical tones, middle C. As almost everyone knows, the frequency of middle C is usually taken as 256 vibrations per second. Using 1,100 feet per second as the speed of sound and substituting these values in the equation,

$$Speed = Wave\ length \times Frequency$$

we get,

$$1{,}100\ feet\ per\ second = Wave\ length \times 256\ vibrations\ per\ second$$

or,

$$Wave\ length = \frac{1{,}100}{256} = 4.3\ feet.$$

This means that when you hear the tone middle C in air at ordinary temperatures the condensations or the rarefactions in that sound wave are a little over 4 feet apart.

Amplitude and Period. — Little needs to be added to what was said in Chapter 28 concerning these two properties. The amplitude of a sound wave is the maximum displacement from its normal position experienced by a particle of the medium as the wave passes. We hear increased amplitude as increased loudness or intensity. The period is the time of one complete vibration of the source. In the case of the middle C tone just mentioned, the period is 1/256 of a second.

These various properties of waves together with their definitions and applications in sound and light are listed in Table 28.

Some Sound Phenomena

In this section we are going to consider four sound phenomena. Everyone is familiar with at least a few of the audible effects caused by the first two, reflection and refraction. Although beats and the Doppler effect produce results that are somewhat less commonly experienced, they are important examples of the way sound waves behave in certain circumstances. As we shall see in the next chapter, all four have their counterparts in the field of light.

Reflection

Reflection of sound, as the name indicates, consists simply in sound waves bouncing off some sort of surface. Echoes and rolling thunder constitute two of the most common reflection phenomena.

TABLE 28. PROPERTIES OF WAVES WITH APPLICATIONS IN SOUND AND LIGHT

Property	*Definition*	*Sound*	*Light*
Speed	Distance traversed per unit time	Approximately 1,100 ft./sec. in air; 15,000 ft./sec. in iron	186,000 mi./sec. Medium — an unsolved problem
Frequency	Rate of vibration of source of waves	Measure of pitch; as frequency increases, pitch gets higher; middle C — 256 vibrations per second; high C — 512 vibrations per second	Measure of color; as frequency increases, color progresses toward violet end of spectrum. Green — 6×10^{14} vibrations per second
Wave Length	Distance between corresponding points on adjacent waves	Measure of pitch; as wave length increases, pitch gets lower; high C about 2.15 ft.; middle C about 4.3 ft.	Measure of color; as wave length increases, color progresses toward red end of spectrum. Green — 20 millionths of an inch
Amplitude	Greatest displacement of a particle of the medium from its normal position	Measure of loudness; as amplitude increases, sound gets louder	Measure of brightness; as amplitude increases, light gets brighter
Period	Time of one complete vibration	For middle C, 1/256 of a second	For green, $\frac{1}{6 \times 10^{14}}$ seconds = 1.7×10^{-15} seconds

In the first of these there has usually been but one reflection, as when one speaks a few words and then hears them repeated a little later after having been reflected from a stone cliff some distance away. Rolling thunder, on the other hand, results from a number of reflections of the original sound from a series of surfaces such as a line of hills. Other common examples of sound reflection will occur to the reader.

Law of Sound Reflection. — The law of reflection, which applies to all kinds of waves, was illustrated for light in Fig. 21 in Chapter 5. A diagram exactly like (a) of that figure could be used to show how sound waves are reflected from a flat or plane surface. Stated formally the law is, *Waves are reflected in such a manner that the angle of reflection equals the angle of incidence*, where those angles are as



FIG. 174. An expanding water wave. (From "College Physics," Arthur L. Foley, The Blakiston Co., 1941.)



FIG. 175. A water wave reflected by a floating beam. (From "College Physics," by Arthur L. Foley, The Blakiston Co., 1941.)

they were defined in the earlier figure to which reference was just made. Obviously, for a plane reflecting surface, the angles between the two rays and the surface itself are also equal.

In the above paragraph we discussed reflection in terms of the direction of travel of the waves. The waves themselves, as pictured earlier in this chapter, consist of a series of alternating condensation and rarefaction wave fronts. Actual wave reflection is pictured in Figs. 174–177. In one case, a water wave is shown as it expands and then as it is reflected from a floating board. In the other, analogous pictures of a sound wave caused by an electric spark are shown.

Sound Reflection in Auditoriums. — Whether or not an auditorium has satisfactory acoustical properties depends largely on its sound reflection characteristics. One of the two principal defects that can be caused by reflection of sound is excess *reverberation*, caused by repeated reflections from the various surfaces in the room. Materials

differ greatly with regard to the percentage of energy in a sound wave that they reflect. Hard plaster reflects some 97 per cent of the wave energy that strikes it, absorbing only 3 per cent. This means that it is a better sound reflector than the best silvered plate glass mirror is a light reflector, because the latter absorbs something like 10 per cent of

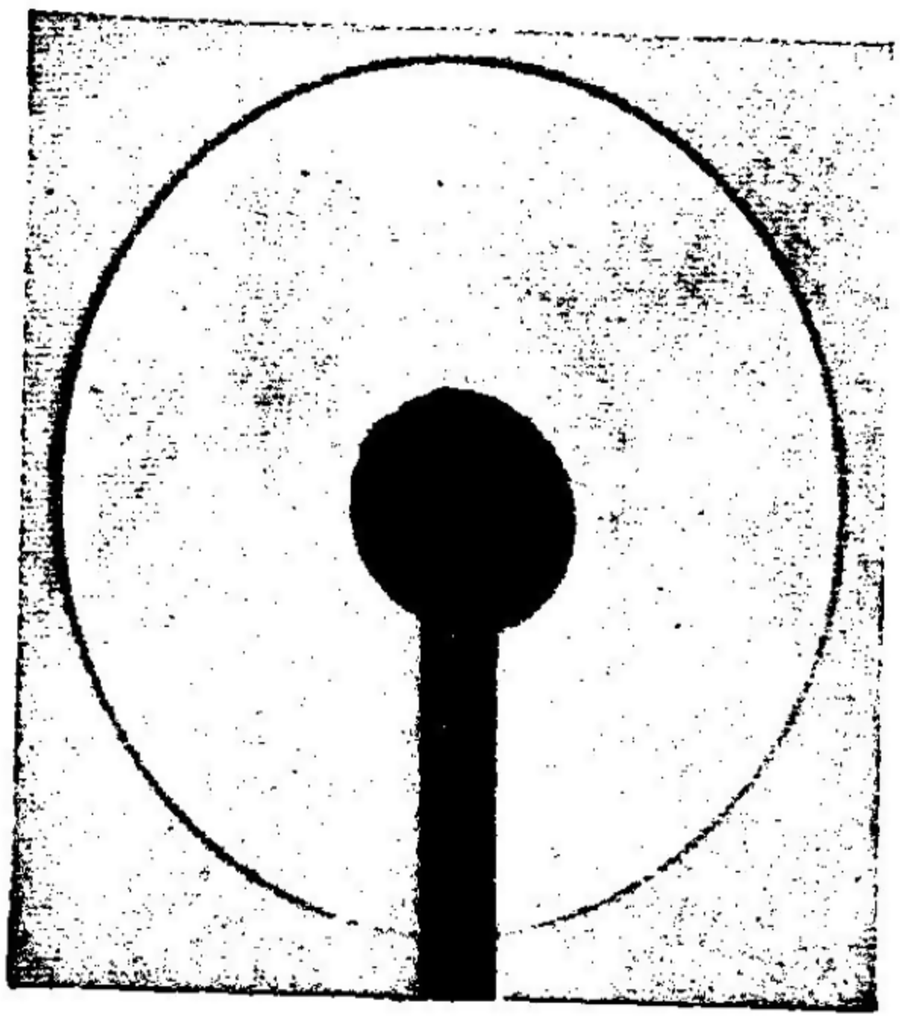


FIG. 176. An expanding sound wave produced by an electric spark (From "College Physics," by Arthur L. Foley, The Blakiston Co., 1941.)

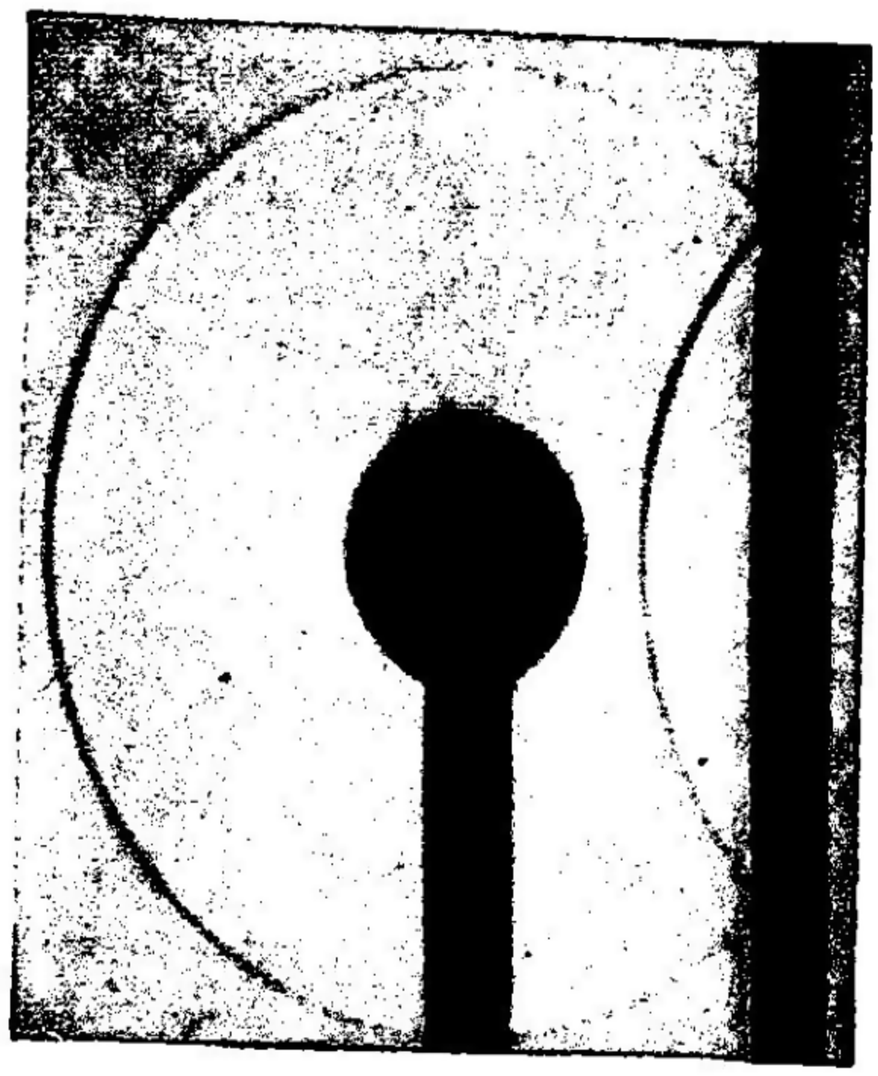


FIG. 177. A sound wave reflected by a glass plate. (From "College Physics," by Arthur L. Foley, The Blakiston Co., 1941.)

the energy in the light that falls on it. The accompanying table gives the percentage of sound energy reflected by various materials.

TABLE 29. REFLECTION PROPERTIES OF A GROUP OF MATERIALS

Substance or Object	*Per Cent of Sound Energy Reflected*
Open window (assumed value)	0
Audience with all seats occupied	10
Heavy curtains	60
Acoustical plaster	30 to 80
Unvarnished wood	94
Ordinary plaster, varnished wood, brick, glass, linoleum	97

Now let us consider a room with hard plaster walls. In it a sound would have 97 per cent of its original energy left after one reflection. After a second reflection there would remain

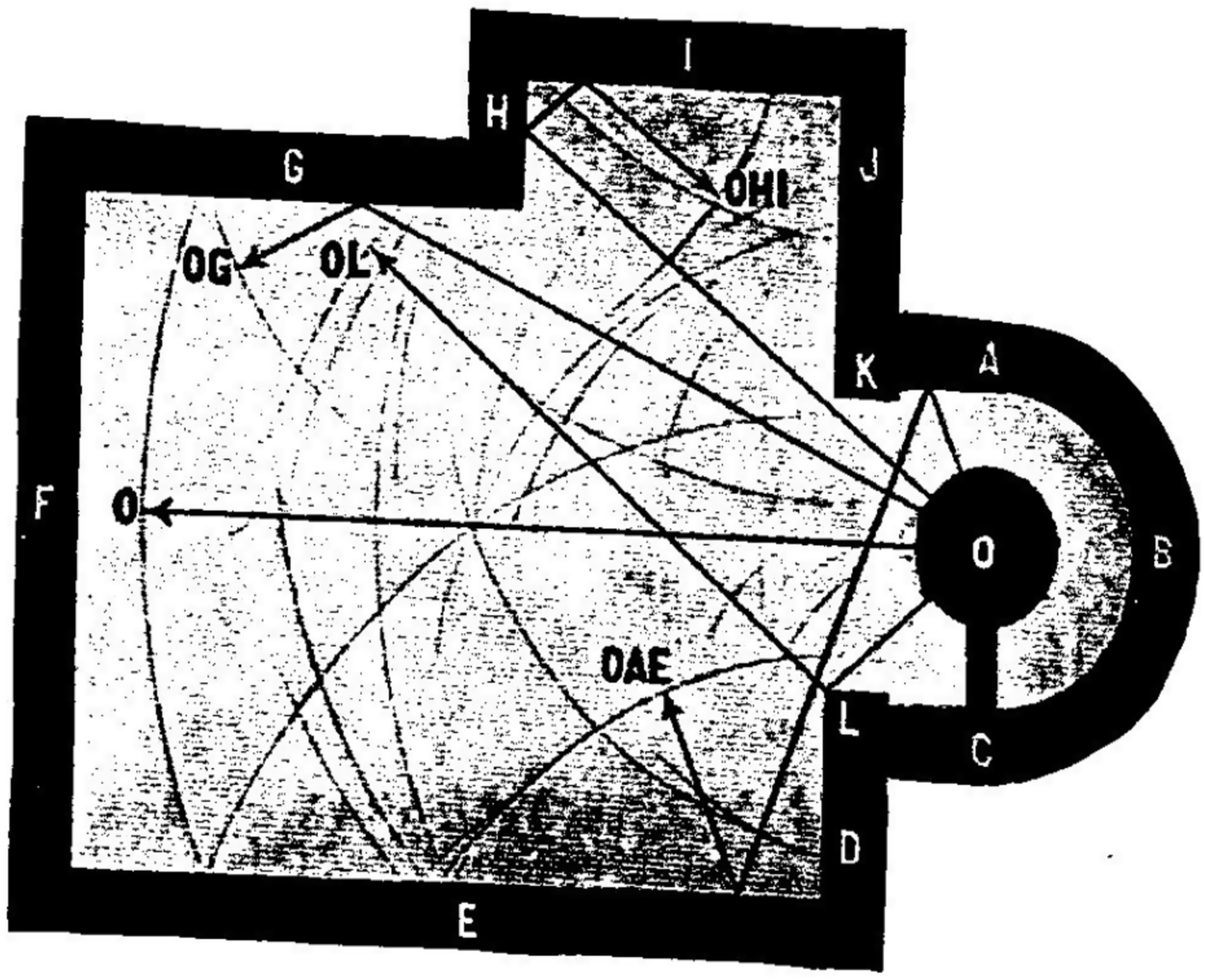
$$97\% - (3\% \text{ of } 97\%) = 95.1\%.$$

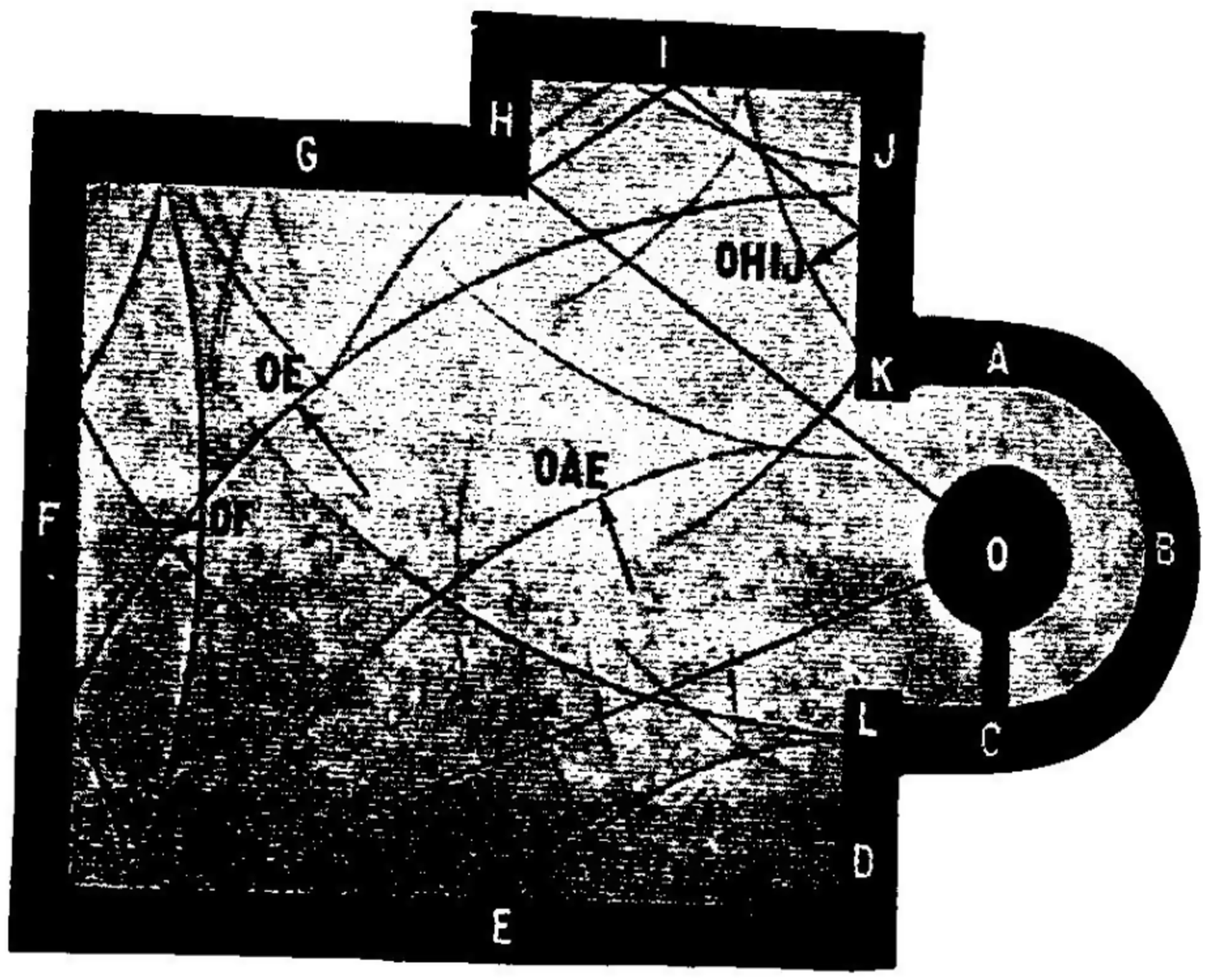
After three reflections,

$$95.1\% - (3\% \text{ of } 95.1\%) = 92.2\%.$$

Hence, 92.2 per cent of the original energy would still be present, and so forth. After twenty such reflections more than half of the original sound energy would be unabsorbed. At this rate the wave could be reflected hundreds of times before being reduced in energy to the point of becoming inaudible. If you, as a member of an audience in such a room, were to hear this sound after perhaps its 300th reflection, it would be mixed up with all the words or music that had been emitted since it originally came from the source. This kind of repeated reflection is pictured in Figs. 178 and 179. The heavy black lines represent directions of travel of waves and the light curved ones the actual wave fronts. These pictures and the explanation under them should be carefully studied by the reader. Reverberation, then, is the conglomeration of sound which results from such multiple reflection. The so-called reverberation period for any given auditorium is the length of time that a sound can be heard after the source stops emitting it. While too much reverberation is extremely unsatisfactory, none at all is almost equally undesirable. In an auditorium with a reverberation period of zero all speech and music would sound flat and lifeless. While there is no reverberation period that is best for all auditoriums, in general the most satisfactory time is between one and two seconds. The optimum time varies somewhat depending on the size and shape of the room and the particular use to which it is to be put.

A second factor that must be considered in auditorium design is the reflection of sound from curved surfaces. In Chapter 5 when we were discussing the reflection telescope we saw how light could be brought to a focus by reflection from a curved mirror (see (b) of Fig. 21). Exactly the same sort of thing can occur with sound waves when they are reflected from curved walls and ceilings. In an auditorium this tends to produce an unequal distribution of sound throughout the room. In extreme cases it may mean that the occupants of seats that are located in spots where waves focus, receive the sound from the stage or rostrum with practically undiminished intensity, while people a few seats away may be in a " dead " spot and receive





FIGS. 178 and 179. Photographs of sound impulses produced by an electric spark behind the disk O on the stage of a plaster paris model of a Chicago auditorium. The letters indicate the paths traversed by the waves. For example the wave O in Fig. 178 has not yet reached the rear wall. It is marked OF in Fig. 179, as it has struck the wall F and been reflected back toward the stage. The wave OHI in Fig. 178 has struck the wall J in Fig. 179 and is therefore marked OHIJ. Note that the wave system from a single pulse and after but one rear wall reflection is rather complicated. Try to imagine what it would be after a score of such reflections, which is a small number, if there were an orchestra on the stage. (From "College Physics," by Arthur L. Foley. The Blakiston Co., 1941.)

almost no reflected sound at all. The so-called " whispering gallery " effect results from focusing of sound in this fashion. Perhaps the most famous example of this phenomenon is found in the Mormon temple in Salt Lake City. A longitudinal cross-section of that auditorium has the shape of half an ellipse, as shown by the solid part of the diagram in Fig. 180. Now an ellipse has two focus points indicated by F and F' and it is a property of the geometry of the figure that waves or particles moving in a straight line from either focus will be reflected at the circumference in such a manner as to pass through the other focal point. For example, on an elliptical billiard table, any ball started from one focus, after reflection from the cushion, would pass through the other one. In the Mormon temple, a very slight sound made at the speaker's desk which is near a focus point, can be heard with considerable clarity at a certain point near the rear of the room, 175 feet away. Readers who have visited the national capitol building will recall having heard the guide demonstrate the whispering gallery effect in the Hall of Statues.

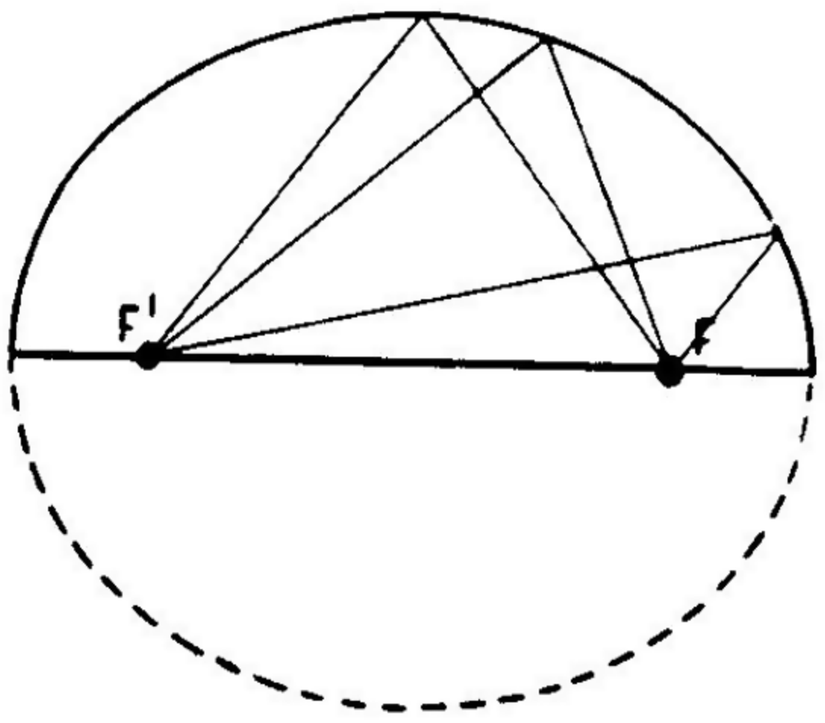


FIG. 180. Diagram illustrating the focussing of sound waves in an auditorium with a cross-section which is half of an ellipse.

Although curved surfaces often produce undesirable acoustical effects, they also can be used to give a more even distribution of sound than would otherwise be possible. It is for this reason that so-called " shells " are often placed back of a stage or orchestra platform, particularly in outdoor amphitheaters.

Refraction

The reader has already been introduced to the fact of refraction as it occurs in light, in connection with the description in Chapter 5 of the refracting or lens-type telescope. It will be recalled that the phenomenon as pictured there was simply one of bending or change of direction of light rays as they passed through the glass lens. Now we are going to ask what causes this bending, discussing it in this chapter in connection with sound waves. The reason for the kind of bending of a wave that we call refraction is the oblique passage of the wave

from one medium into another in which it travels at a different speed. This effect can be illustrated by the example shown in Fig. 181, one that has been employed in countless elementary physics books. Part (*a*) shows several columns of soldiers marching on a concrete parade ground. They are moving in the direction indicated by the arrow and are approaching a sandy field, obliquely. Other things being equal,

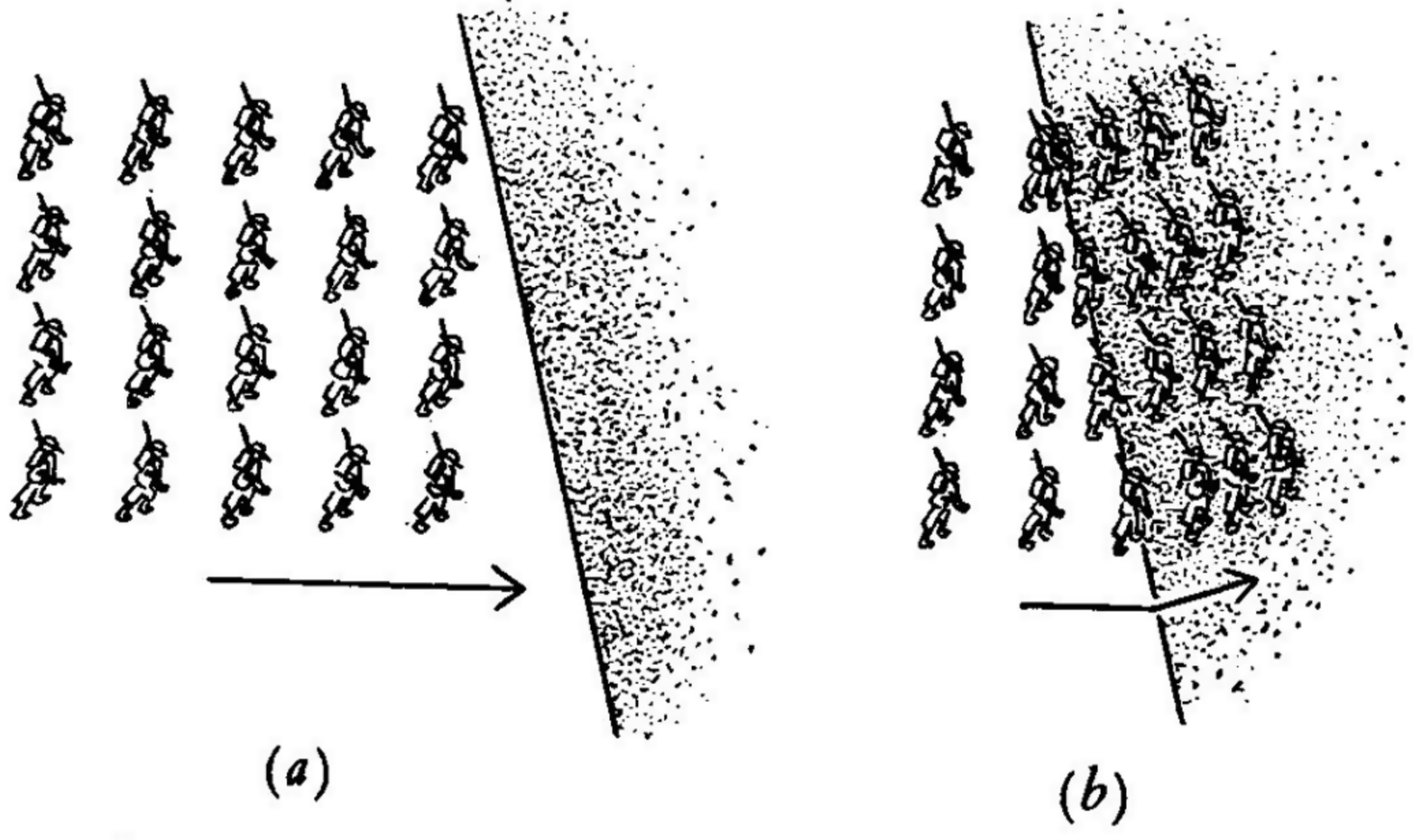


FIG. 181. Diagram illustrating refraction of waves.

their rate of march in the sand will be less than it is on the concrete. In other words, they are about to move from one medium into another in which their velocity will be less. Part (*b*) shows the result of this change in medium. The end of each column that gets into the sand first is slowed down while the other end is still marching at its original speed on the concrete. The effect is to swing the column toward the sand. As soon as all the men are in the sand they proceed in a straight line, but in a direction somewhat different from their original one. Now suppose instead of columns of soldiers going from concrete into sand, we have sound waves moving obliquely from air into, say, rubber. The part that strikes the rubber first will be slowed down first and exactly the same sort of bending or refraction will occur. In fact, sound can be focused with a rubber lens of the same shape as the glass refracting telescope lens that was described in Chapter 5.

Everyone has noticed that music or speech coming over water on a cool evening often seems unnaturally loud. This is because of atmospheric refraction of sound. Frequently, on such evenings, the temperature of the air at different levels is as shown in Fig. 182, that is, cold air near the surface and warmer air higher up. It will be re-

called that sound travels faster in warm air than in cold. So the wave fronts of sound as they come from the bugle at the left do not move out as spherical surfaces as they would do in a uniform medium. Instead, parts of the front that get into the warmer air speed up with respect to the portions near the water and are bent back down again as shown. Therefore, an unusually large proportion of the sound energy reaches the opposite shore, which means the bugle tones sound louder than usual at that point. If the stratification of the air is reversed with the cold layer at the top, then the wave fronts that would normally move directly across the water will be bent up and it may be difficult to hear the bugle at all on the right-hand shore.

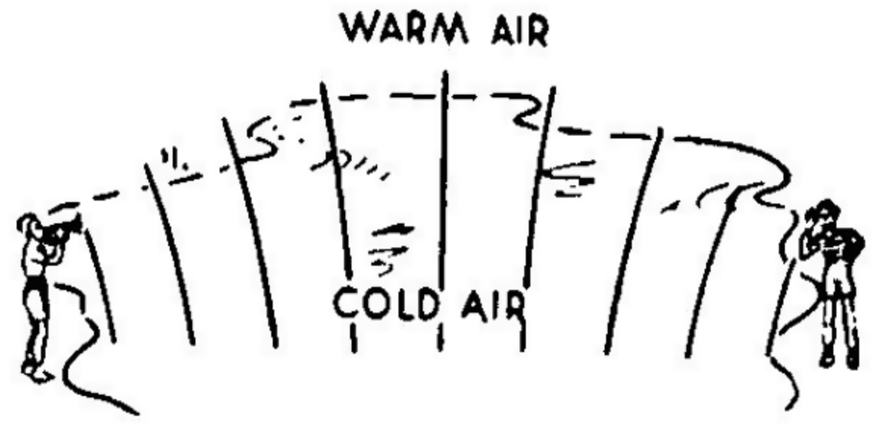


FIG. 182. Diagram illustrating refraction of sound in air which is colder at the earth's surface than it is a few feet higher.

Beats

The phenomenon of beats occurs when two notes of slightly different frequencies are sounded simultaneously. One hears it as a tone which varies regularly and continuously between loud and soft, that is, as a sort of tremolo. Let us approach an explanation of the effect by way of the following illustration. Suppose a man who takes steps 3 feet long and a boy whose steps are only 2½ feet long start out in step and walk side by side. By " in step " we mean, of course, that they start off on the same foot. When they have gone a distance of 30 feet, the man has taken exactly 10 steps and the boy 12. Since each has taken an even number of steps they are in step again at the end of the 30-foot trip. Now if they start in step and finish in step, but one takes more steps than the other, then there must have been one or more places along the way when they were exactly out of step. In this case there was one such place, at the end of 15 feet. At that point the man had taken five steps and the boy six, putting them on opposite feet or exactly out of step. Let us translate this over into the wave situation pictured in Fig. 183. Here we have two waves, one with a frequency of 5 vibrations per second and the other with 6. Assuming that they were sounded simultaneously, note that they start out in step, that is,

both start with a condensation. In wave language we would say they are "in phase." They are in phase again at the end of the interval shown, both ending with a rarefaction. Note, however, that at a point halfway along they are exactly out of phase, that is, a condensation in one wave is opposite a rarefaction in another. Recalling that a

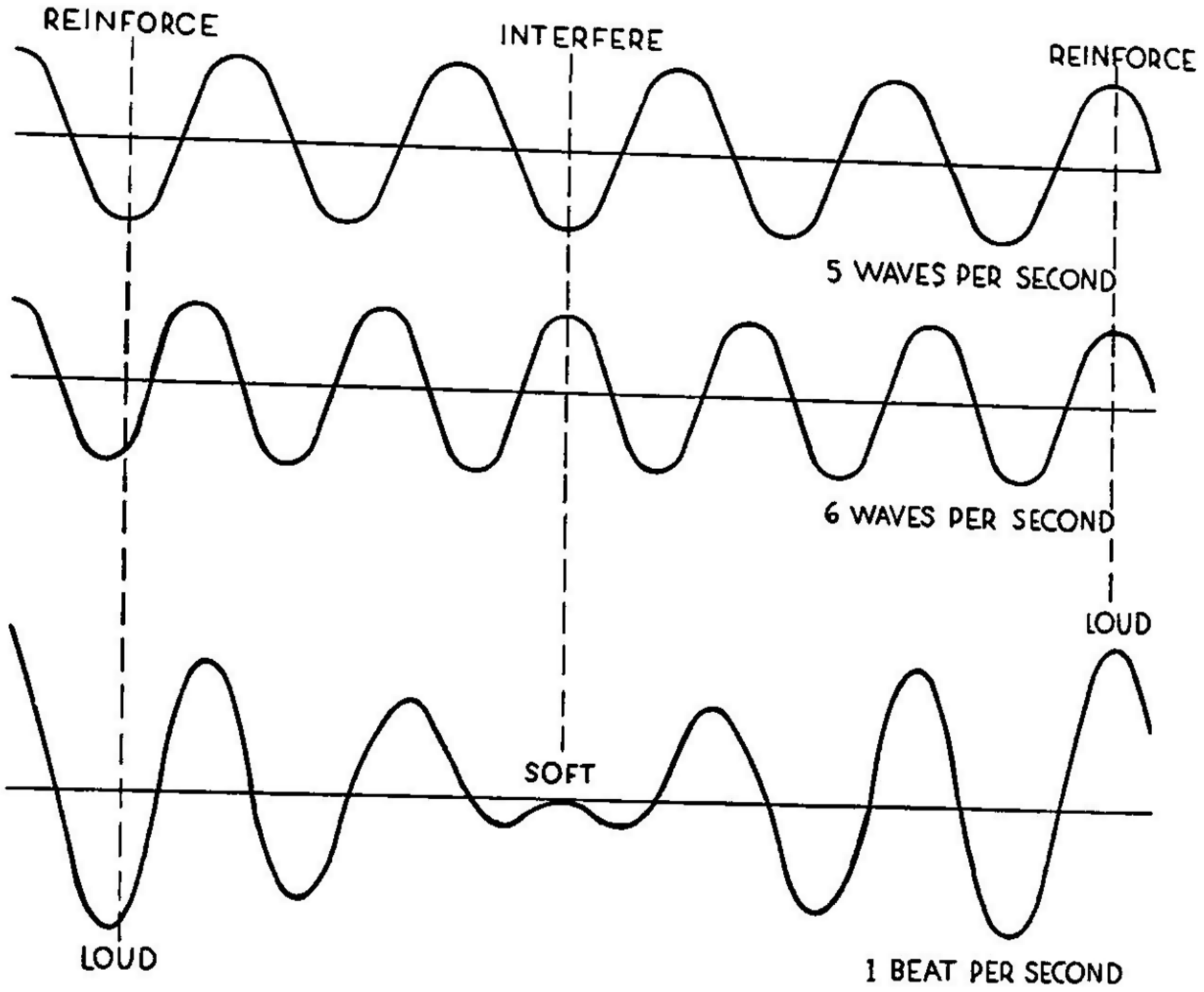


FIG. 183. Diagram showing how two tones of slightly different frequencies combine to produce the phenomenon of beats.

condensation is a compressed region and a rarefaction a partial vacuum, then a combination of the two should produce an approximately normal distribution of air particles, that is, approximate silence. At such a point the waves are said to interfere. At the ends, however, where the waves are in phase, the condensations and rarefactions reinforce each other, increasing the loudness of the sound. Consequently, the combination tone that one hears is first loud, then soft, then loud again, and so on. This resultant or combination tone is represented at the bottom of the diagram. There will be as many beats per second as the difference in the frequency of the two tones. For example, tuning forks of 256 and 258 vibrations per second respectively would produce 2 beats per second when sounded together. The phenomenon can easily be demonstrated by taking two identical tuning forks and

loading one with a little wax or a rubber band in order to slow it down slightly.

Beats are made use of in tuning pianos and other musical instruments. If it is desired to bring two strings to exactly the same pitch, the tension of one is changed until all beats disappear. Tremolo effects in pipe organs are produced by having two pipes of slightly different frequencies sound at the same time.

The Doppler Effect

The Doppler effect, as it occurs in light, was mentioned in Chapter 5 in connection with the measurement of star velocities. It will be recalled that, as pictured there, the phenomenon concerned a small shift in frequency, that is, a change in color of the light emitted by a star because of its motion either toward or away from the earth. Since variations in frequency in sound are heard as changes in pitch, it is clear that the Doppler effect in this field should concern shifts in pitch as a result of the motion of the sounding source. Probably the most commonly heard example of the effect occurs when an automobile approaches with its horn blowing. At the instant the car passes, one hears a sudden and quite appreciable drop in the pitch of the sound. Or if one is standing on a railway platform a distinct lowering of the pitch of the whistle is noticed as a train goes by.

As to the explanation of the phenomenon, again we will make our approach by way of an analogy. Suppose two men, A and B, are standing some distance apart and A is throwing baseballs at B at the rate of 20 per minute. As long as both men stand still it is obvious that B receives 20 baseballs per minute. But now suppose A runs toward B, still throwing balls at the same rate. Since each baseball has a little shorter distance to go than the one that preceded it, B receives more than 20 per minute. On the other hand, if A runs away from B, throwing the balls back at him still at the same rate, they have successively farther to go and B receives fewer than 20 per minute.

Now let us substitute a middle C tuning fork for A. It is throwing 256 sound waves per second at B. As long as both are stationary, B receives 256 per second and says that he hears middle C. But if the fork moves rapidly toward him, more than 256 waves reach his ears per second and he hears a higher tone. Conversely, if the fork is moving away from him he hears a tone lower than the one actually

being emitted. How much higher or lower the tone is depends on how fast the source is moving. The effects will be similar if the source remains fixed and the observer moves.

In the case of the automobile horn mentioned above, as the car approaches one hears a tone of higher pitch than the normal one. The instant it passes that pitch drops to one lower than the one that is heard by the driver of the car who is not moving with respect to the

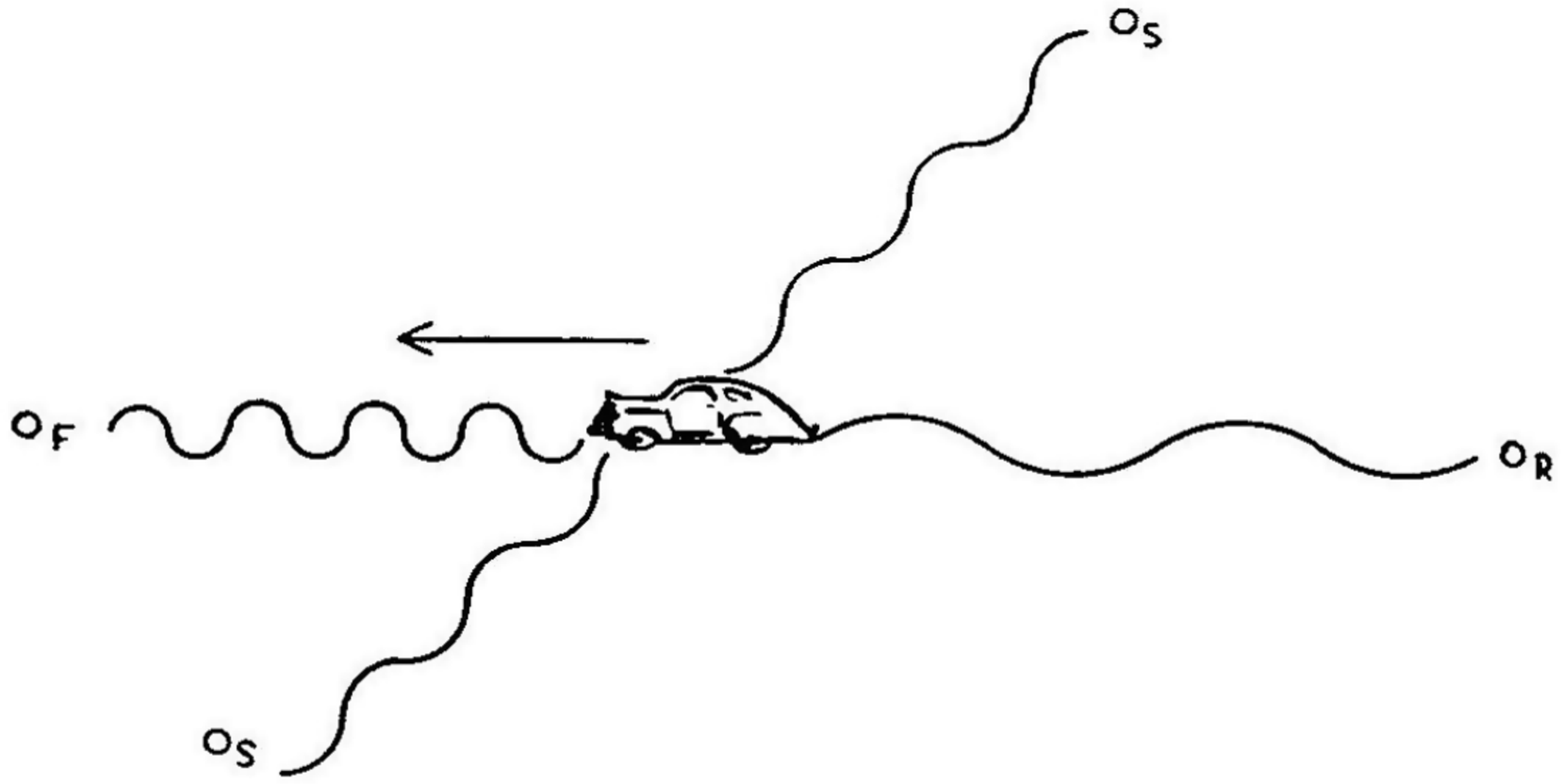


FIG. 184. Diagram illustrating Doppler Effect in sound. When the automobile horn is sounded, observers at O_S hear the true pitch; those at O_F hear a tone higher than the true one; while those at O_R hear a lower one.

horn. Note that the shift is instantaneous. The pitch does not gradually change from a higher one to a lower. The diagram shown in Fig. 184 illustrates how the horn would sound to observers located at four different positions with respect to the car. The two at the sides of the car (O_S) hear the horn's true pitch. The one in front (O_F) hears a higher pitch and the one in the rear (O_R) a lower one.

THE PHYSICAL BASIS OF MUSIC

The dictionary defines music as " the art or science of pleasing, expressive, intelligent combinations of tones." It is our purpose in the rest of this chapter to consider some of the physical principles that are fundamental to the production of such tone combinations.

Characteristics of Musical Tones

Musical tones are characterized by three properties, two of which we have already discussed in some detail. The first of these, *pitch*, has already been described as being determined by the frequency of the

vibrating source. The chart shown in Fig. 185 gives the range of frequencies characteristic of various musical instruments and voices. Note that the lowest note on the piano has a frequency less than 15 vibrations per second above the lower limit of audibility.

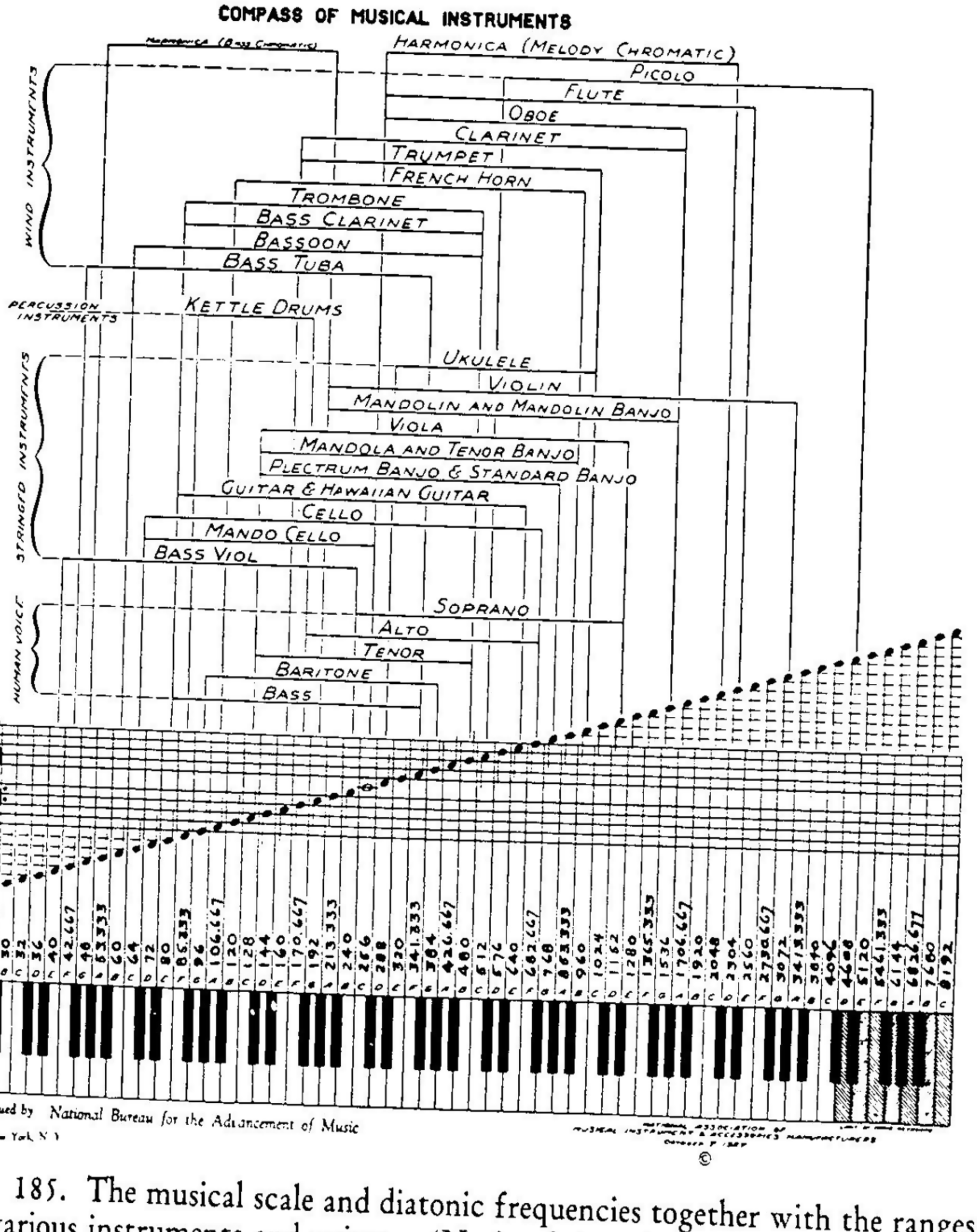


FIG. 185. The musical scale and diatonic frequencies together with the ranges of various instruments and voices. (National Association of Musical Instrument and Accessories Manufacture.)

The second characteristic, *intensity*, depends on the amplitude of the sound waves. Tones may range all the way from those so low as to be barely audible to those loud enough to be painful. Experiments show that these limits vary with frequency as well as with individuals.

It is obvious that pitch and intensity do not entirely describe a musical tone. For example, middle C played on a saxophone does not

sound exactly the same as middle C sung by someone, or played on a violin or on some other instrument, even though the intensity and pitch may be the same in all cases. The very fact that we can distinguish between instruments in this fashion makes it clear that tones possess a third characteristic. This is called *quality* or timbre, and is determined by what are known as overtones. We will make use of the series of vibrating strings pictured in Fig. 186 to explain just what

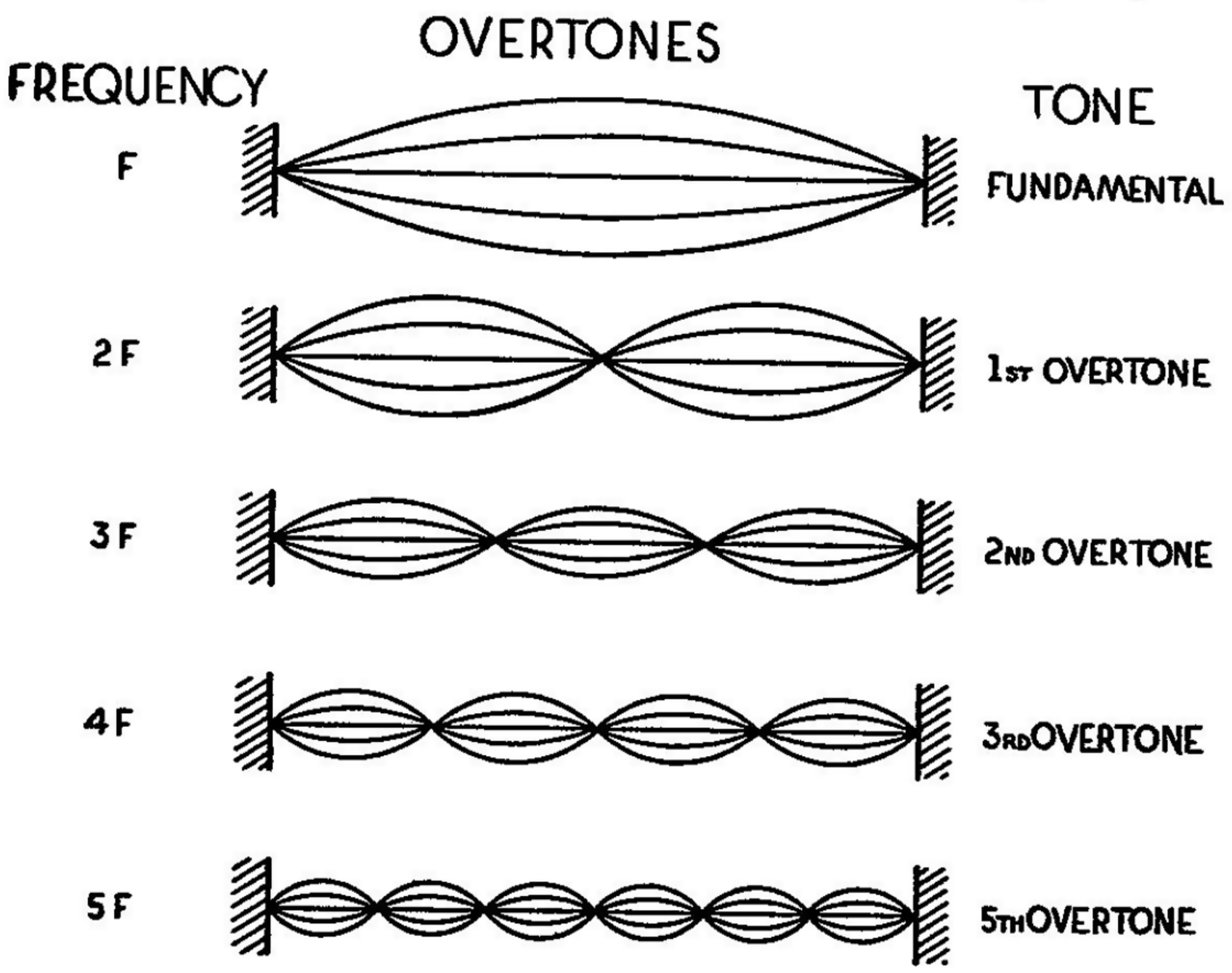


FIG. 186. Diagram showing relationship of modes of vibration to frequencies and overtones for a vibrating string. (After White.)

we mean by overtones. The lowest possible tone that a vibrating string can emit under given conditions of length and tension is the one emitted when it is vibrating in one piece, as shown by the top string in the figure. This is its fundamental tone. However, it is possible to make it vibrate in halves as shown by the second string from the top. When it does, the tone it emits has twice the frequency of the fundamental and is called the first overtone. The relation of other numbers of parts in which the string may vibrate, to the overtones and their frequencies will be clear from the rest of the figure. Since the term " harmonic " is sometimes used instead of overtone it should be pointed out that the fundamental is the first harmonic, the first overtone is the second harmonic, etc.

Now the important point, as far as the quality of the tone emitted

by a vibrating string is concerned, is the fact that it may vibrate in two or more of these ways at the same time. Fig. 187 shows a string vibrating in both one and three parts. Such a string would be emitting a combination of the fundamental and the second overtone. The tone that one would hear from this string would sound a little different from that emitted by a string vibrating in, say, one and two parts at the same time, although both would have the same fundamental frequency. In other words, one would have a different quality than the other. Although we have used vibrating strings for our illustration, the same thing applies to vibrating plates, air columns, rods, and bells. This means that the quality of the tone coming from any musical instrument is determined by the number, selection, and intensity of the overtones that are present. There are devices such as the phonodeik and oscillograph which make it possible to photograph these complex combinations of the fundamental and one or more overtones emitted by various instruments. The curves shown in Fig. 188 were drawn

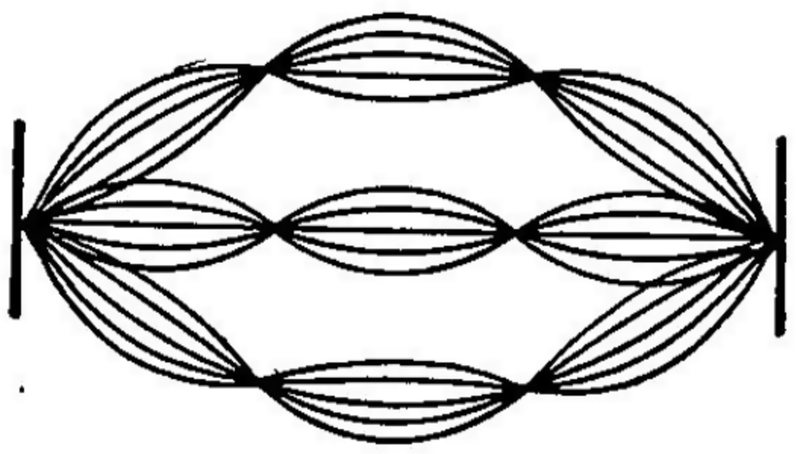
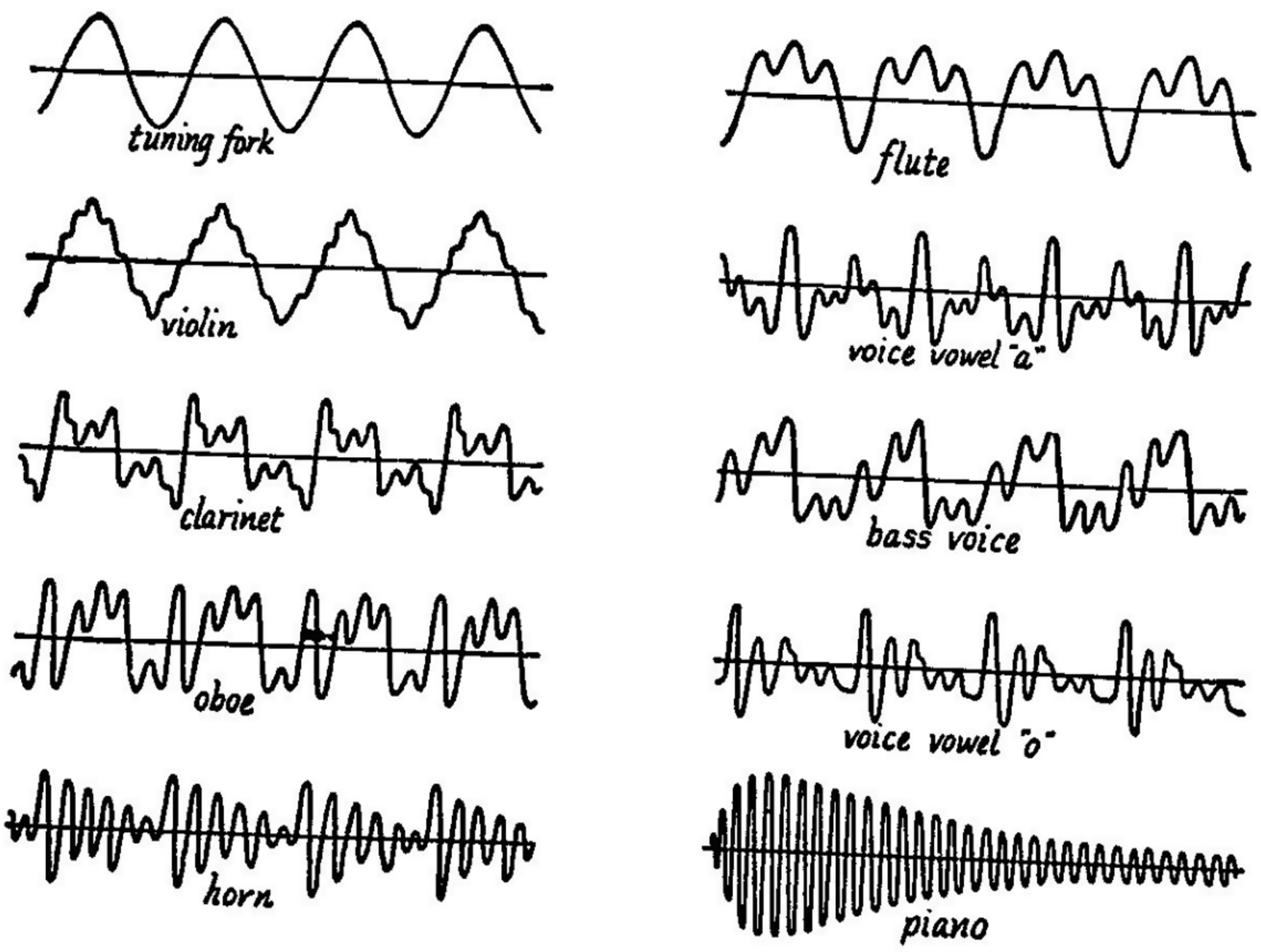


FIG. 187. A vibrating string emitting its fundamental tone and 2nd overtone simultaneously.



IG. 188. Types of waves emitted by various instruments and voices. (From Classical and Modern Physics," by Harvey E. White, D. Van Nostrand Co., 940.)

from such photographs made by the late Dr. Dayton C. Miller, one of the world's outstanding authorities in this field. Note that all of the tones except the one from the piano have the same fundamental frequency, yet because of the particular set of overtones present, vary greatly in details. Note also that only the tuning fork gives the simple fundamental tone with no overtones present.

The Interval Law

Everyone is familiar with the fact that certain combinations of tones produce harmony when sounded together, while others are discordant. The interval law is an experimentally determined relationship by which one can predict, from their frequencies, whether any two or more particular tones will sound well or poorly together. Before stating the law, let us describe an experiment that we might perform to discover at least an approximation of it for ourselves.

The purpose of the experiment would be to see if we could find any relationship among the frequencies of tones that would give us a basis for deciding which ones would produce harmony when sounded together and which ones discord. The apparatus for a rather limited trial could be the eight white keys on a piano, running from middle C to the next C above. The procedure would be to sound these tones in all sorts of combinations of pairs, noting in each case what the frequencies of the two tones were and whether or not they sounded harmonious. The frequencies of these eight tones, on the basis of 256 vibrations per second for middle C, are given in part (A) of Table 30. Some of the data we would get by performing such an experiment appears in part (B) of the same table. The fractions given there were determined by reducing to its lowest terms the fraction one would get by using one of the pair of frequencies as the numerator and the other as the denominator. In the first one C = 256 and D = 288. Therefore,

$$\frac{C}{D} = \frac{256}{288}, \text{ which reduced to its lowest terms} = \frac{8}{9}.$$

This simply says that the ratio between the frequencies of C and D is 8/9 or 8 to 9. If the experiment were performed the combinations shown in the table in boldface type would sound well together while

the others would produce discord. Readers who have access to a piano would do well to check the combinations listed in the table and as many additional ones as possible.

TABLE 30. FREQUENCIES OF OCTAVE BEGINNING WITH MIDDLE C AND RATIOS OF FREQUENCIES OF VARIOUS PAIRS OF TONES. THOSE IN BOLDFACE ARE HARMONIOUS COMBINATIONS.

A. Frequencies in vibrations per second

C — 256
D — 288
E — 320
F — $341\frac{1}{3}$
G — 384
A — $426\frac{2}{3}$
B — 480
C′ — 512

B. Combination Ratios

$\frac{C}{D} = \frac{8}{9}$ $\frac{C}{E} = \frac{\mathbf{4}}{\mathbf{5}}$ $\frac{C}{F} = \frac{\mathbf{3}}{\mathbf{4}}$

$\frac{C}{G} = \frac{\mathbf{2}}{\mathbf{3}}$ $\frac{C}{A} = \frac{\mathbf{3}}{\mathbf{5}}$ $\frac{C}{B} = \frac{8}{15}$

$\frac{C}{C'} = \frac{\mathbf{1}}{\mathbf{2}}$ $\frac{E}{G} = \frac{\mathbf{5}}{\mathbf{6}}$ $\frac{D}{F} = \frac{27}{32}$

$\frac{B}{C'} = \frac{15}{16}$ $\frac{E}{C} = \frac{5}{8}$ $\frac{G}{A} = \frac{9}{10}$

In analyzing even the small amount of data given above, one sees that the ratios for the harmonious combinations are smaller numbers than those for the discordant combinations. By taking many more data we would eventually be able to determine at just what ratio point harmony leaves off and discord begins. The results of such investigations lead to the *interval law* which may be stated,

The ear recognizes as harmonious only combinations of tones whose frequencies are in the ratio of whole numbers of six or less.

It should be recognized that the division point between six and seven was chosen purely on the basis of physiological reaction. That is, ratios of six or less sound well to most occidental peoples, while combinations with ratios greater than six do not.

In musical terminology the ratio between the frequencies of two tones is called an *interval*. Certain of the musical intervals, which are given specific names, are listed in Table 31.

The Major Diatonic Scale

In the sequence of tones that we call the major diatonic scale, the intervals are strictly in accordance with the values given above. The scale is based on three major triads. A major triad, as listed in the

TABLE 31. SOME MUSICAL INTERVALS

Interval	*Frequency Ratio*	*Example*
Octave	1 to 2	C–C′
Third	4 to 5	C–E
Fourth	3 to 4	C–F
Fifth	2 to 3	C–G
Sixth	3 to 5	C–A
Major triad	4 to 5 to 6	C–E–G
Half tone	15 to 16	E–F

table, consists of three tones whose frequencies are in the ratio of 4 to 5 to 6. The familiar *do-mi-sol* of any key is an example. Now, using a *do* with a frequency of 24 vibrations per second, we shall construct a major scale. First the steps will be listed in tabular form, and then each one discussed briefly.

	do	*re*	*mi*	*fa*	*sol*	*la*	*ti*	*do′*	*re′*
Step 1	24		30		36				
Step 2					36		45		54
Step 3	24	27						48	54
Step 4				32		40		48	
Complete	24	27	30	32	36	40	45	48	54

Step 1. If *do* = 24 vibrations per second and *do-mi-sol* is to be a major chord, then *mi* = 5/4 of 24, or 30, and *sol* = 6/4 of 24, or 36.

Step 2. The sequence *sol-ti-re′* constitutes the second major chord. Applying the procedure of Step 1, we get *ti* = 45 and *re′* = 54. Readers with some musical background will recognize that *if* the above scale were the key of C, then what we have called *sol-ti-re′* would be *do-mi-sol* in the key of G. Thus, this second major chord is also a *do-mi-sol* sequence.

Step 3. Now we make use of the octave relationship in which the frequency ratio is 1 to 2. Thus, if lower *do* = 24, the next higher *do′* = 48; and if *re′* = 54, then lower *re* = 27.

Step 4. The third major chord is composed of the tones *fa-la-do′*, of which we know *do′* = 48. Consequently, *la* = 5/6 of 48, or 40 and *fa* = 4/6 of 48, or 32. Here again, as in Step 2, we actually have a *do-mi-sol* in another key. This time it would be the key of F *if* our example were the key of C.

In this way we have determined all the frequencies for a major diatonic scale for which *do* = 24 vibrations per second. Since the intervals are all correct, a scale starting with any other value of *do* could be obtained by multiplying all the above values by some particular number. For example, the reader will find that if he multiplies all of the above frequencies by the fraction $\frac{32}{3}$, he will get the values given in part (A) of Table 30 for the scale starting with *do* = 256 vibrations per second.

The Tempered Scale

Although the diatonic scale has exact intervals which are simple whole numbers, it presents certain practical difficulties from the standpoint of instrumental music. These difficulties arise out of the fact that it is not composed of equal interval steps. Instead they are as follows:

do	re	mi	fa	sol	la	ti	do
	$\frac{8}{9}$	$\frac{9}{10}$	$\frac{15}{16}$	$\frac{8}{9}$	$\frac{9}{10}$	$\frac{8}{9}$	$\frac{15}{16}$

As a result of this, an instrument that was tuned to exact diatonic frequencies for a certain key could be played only in that key. This, of course, would be particularly undesirable in instruments of fixed pitch, like pianos and organs. Furthermore, to play such a piano even in the key to which it was tuned, one would need two black keys every place that ordinarily there is one, because in strict diatonic intervals the sharp of a given tone has a different frequency from the flat of the next higher tone.

However, the human ear does not readily detect small frequency differences and so it has been possible to construct a so-called *tempered scale* in which all the intervals are the same. The frequencies of this scale are shown in Table 32 alongside the true diatonic values.

Note in this table that in no case does the tempered-scale value differ by more than a few frequency numbers from the diatonic figure. It will be seen that there are 12 half-tone intervals in the complete octave. If these are all to be equal, then the value of one interval must be the number that multiplied together 12 times will equal 2,

TABLE 32. DIATONIC AND TEMPERED SCALE FREQUENCIES

Name in Key of C	*Note*	*Diatonic Frequencies*	*Tempered Frequencies*
do	C	256	256
	C#, Db	——	271.2
re	D	288	287.3
	D#, Eb		304.4
mi	E	320	322.5
fa	F	341⅓	341.7
	F#, Gb		362.0
sol	G	384	383.6
	G#, Ab		406.4
la	A	426⅔	430.5
	A#, Bb		456.1
ti	B	480	483.2
do′	C′	512	512.

since the final frequency is to be just twice the beginning one. In other words, in mathematical terminology, the value of an interval in the tempered scale is the 12th root of 2, which is approximately 1.059. If the reader checks this he will find that,

256 × 1.059 = 271.2, the frequency of C#,
271.2 × 1.059 = 287.3, the frequency of D,
287.3 × 1.059 = 304.4, the frequency of D#, . . .
and so forth.

Types of Pitch

To avoid confusion in our discussion of the basis of the major scale we have thus far used exclusively a middle C with a frequency of 256 vibrations per second. However, it should now be clear from the manner in which the major scale is built up that it is the intervals or ratios and not the specific frequencies that are important. Using the technique applied above one could construct a perfectly good major scale using as *do* a middle C of 200, 260, 300, or any other number of vibrations per second. It happens that the musical profession uses the A above middle C, instead of middle C itself for the defining frequency. In the scale we have used, A has a frequency of 426⅔ vibrations per second. At different times in history, A's of a variety of values have been employed. Some early church music gave

it a value as low as 374. A tuning fork used by Handel was 422.5. During the period from about 1625 to 1825, 420 was the average accepted value. In so-called International pitch, the frequency of A is 435. In order to obtain uniformity in this respect the American musical industry in 1925 agreed on a standard pitch defined by an A with a frequency of 440 vibrations per second. Such a scale gives middle C a frequency of 261.6 vibrations per second.

Questions

Discussion

1. In the light of the statement in this chapter about the dependence of the speed of sound on the elasticity and density of the medium, combined with your knowledge of the kinetic theory, explain why the speed of sound both in a free gas like the atmosphere and in an enclosed gas should increase with temperature.

2. Calculate the approximate number of octaves in the entire range of frequencies audible to the human ear.

3. Would a sound wave with a wave length of 100 feet traveling through air be audible to you? Explain.

4. Calculate the approximate wave length of the highest pitch tone that you could hear.

5. Would two tones sounded together produce harmony if the period of one is three times the period of the other? Explain.

6. Draw a diagram to illustrate what is meant by angles of incidence and reflection.

7. If possible find some local auditorium that was either designed for good acoustics or altered to improve the acoustics and find out the details of what was done.

8. After studying the illustration used in this chapter in the discussion of refraction, draw diagrams to show what would happen to columns of marching soldiers who, (a) moved obliquely from a sandy field onto a concrete parade ground, (b) encountered in the center of a concrete area a circular sandy spot having a diameter about equal to the width of the column.

9. Suppose strings X and Y produce 5 beats per second when sounded together; which one originally had the lowest frequency if slightly raising the pitch of Y reduces the number of beats to 3 per second?

10. Relate the Doppler effect to the way a racing automobile on an oval track sounds to someone in the grandstand.

11. Define pitch, intensity, and quality of musical tones.

12. Using the frequencies given in Table 30 determine whether the following pairs of tones would produce harmony or discord when sounded together: (a) D and E; (b) A and C′; (c) G and C′; (d) D and G.

13. Determine whether or not the tone sequence *re-fa-la* in the major diatonic scale constitutes a major chord.

14. Explain any advantages that the tempered scale has over the diatonic scale.

15. Calculate the following: (a) the frequency of *re* in a diatonic scale for which *do* = 600 vibrations per second; (b) the interval between a frequency of 400 vibrations per second and one of 500.

16. Calculate the frequencies in the tempered scale of the sharp and flat of a tone which has a frequency of 450 vibrations per second.

Multiple Choice

1. You are riding on the rear platform of a train as it approaches a station; if the engineer blows the whistle, which of the following (one or more) are true? (a) The tone you hear is the one actually sounded by the whistle. (b) The tone you hear is lower than the one heard by a man on the station platform. (c) Neither you, the engineer, nor the man at the station hear the true tone. (d) The tone you hear is halfway between those heard by the other two men. (e) All three of you hear the same tone.

2. Which one of the following sound phenomena can only occur when two waves of different frequencies come together? (a) Reflection. (b) Refraction. (c) Doppler effect. (d) Beats. (e) Focusing.

3. List *all* of the following groups of tones that should produce harmony if sounded together, the numbers being frequencies. (a) 222 and 444. (b) 256 and 576. (c) 800, 1,000, 1,200. (d) 111, 222, 333. (e) 100, 110, and 120.

4. Which two of the following are the frequencies of the end tones of a major chord, the middle tone of which has a frequency of 600 vibrations per second? (a) 320. (b) 400. (c) 480. (d) 600. (e) 700. (f) 720. (g) 840.

5. On a day when sound travels 1,080 feet per second, the tone *mi* of the next scale above one in which *do* has a frequency of 400 vibrations per second, has a wave length of between — (a) 1 and 2, (b) 2 and 3, (c) 3 and 4, (d) 4 and 5, (e) 5 and 6 — feet.

True–False

1. The continuous roll of thunder is a refraction phenomenon.

2. In order just to be heard a sound of low frequency must have a greater amplitude than one of intermediate frequency.

3. Any three tones that produce harmony when sounded together may be said to constitute a major chord.

4. In the tempered scale, all half-tone intervals have the same value.

5. The musical interval between two tones exactly two octaves apart is 1/4.

6. All audible frequencies of sound travel at about the same velocity in air of a given temperature.

7. The second overtone of middle A in a scale of standard pitch would have a frequency of 880 vibrations per second.

8. Tones whose frequencies are in the ratio of 3.76 to 5.37 should produce harmony when sounded together.

9. The musical interval between two tones is calculated by subtracting the frequency of one from that of the other.

10. In an auditorium containing hard wooden seats, the reverberation period would be less when the room was full of people than when it was empty.

Chapter 30

LIGHT

In all the references that were made to light in the preceding two chapters we assumed it to be a wave motion. And, indeed it is true that essentially all of our common everyday experiences with this particular kind of energy can be understood or accounted for satisfactorily by thinking of it in this fashion. Consequently, in the first part of this chapter, in which we shall discuss a group of the more familiar light phenomena, we shall continue to think of it as a type of wave motion. Nevertheless, even a survey of the field, if it is to be a fair one, must take some account of a certain group of well-established experimental facts which seem to contradict this wave picture. Therefore, in the latter part of the chapter we shall look briefly at the history of man's ideas concerning the nature of light and see something of where science stands today on the question. It may be mentioned here that the alternative to considering light as a wave motion is to think of it as consisting of tiny particles or bullets of some sort. It might be well for the reader to ask himself with regard to each of the phenomena discussed, " Does it argue for waves, or particles, or isn't it an argument either way? "

Some Light Phenomena

Most of the following manifestations have already been discussed as they occur in sound. It will aid the reader in remembering both fields if he will compare them as we consider the same phenomena in light.

Speed

Everyone today accepts the idea that light requires a definite amount of time to travel from one place to another, that is, it has a finite speed. Also, a fair percentage of people are familiar with the

fact that its speed amounts to about 186,000 miles per second. The enormous magnitude of this figure made it impossible for early scientists to measure it, and so for a long time light was considered by many of them to be propagated instantaneously from one point to another. Aristotle was one who accepted this view.

The first actual attempt to measure the speed of light was made by Galileo. He and an assistant, each equipped with a lantern, stationed themselves on hilltops a mile apart and carried out the following experiment. Galileo, noting the exact time, uncovered his lantern. The assistant uncovered his lantern at the instant he saw the flash from the one held by Galileo. When Galileo saw the light from this second lantern he again noted the time. He reasoned that the interval between his two readings would be the time required for light to travel from him to his assistant and back again, that is, two miles. The experiment was a failure because in none of the trials was Galileo able to detect any time interval at all. The reason it failed, of course, was simply that the experimental technique was too crude to cope with the enormous speed the investigators were trying to measure. The time required for either man to uncover a lantern, or for Galileo to shift his gaze from the lantern to a timepiece would be many times the approximately one one-hundred thousandth of a second it takes light to travel a distance of two miles. The incident is an excellent example of a scientific experiment that was based on perfectly sound theory but failed because of practical performance difficulties. In fact, some of the more recent successful determinations of the speed of light have been based on essentially the same principle.

The first successful determination came as a kind of by-product of some astronomical experiments made about 1676 by the Danish astronomer Olaf Römer (1644–1710). Römer was engaged in a study of certain of Jupiter's moons, particularly one with the comparatively short revolution period of 42 hours. Since Jupiter requires 12 times as long as the earth to complete a trip around the sun, each year we catch up to and pass this largest member of the solar family. Now Römer could observe for a short time the frequency with which the Jovian satellite disappeared behind the planet, and then make out a schedule of just when these eclipses should occur in the future. But when he based such a schedule on his observations made when the earth and

Jupiter were on the same side of the sun, he discovered that 6 months later when the two planets were on opposite sides, an eclipse that was scheduled to occur at midnight did not occur until about 16⅔ minutes after midnight. Then, if he made out a new schedule based on these later observations, after another 6 months had passed and the earth was back on the same side of the sun as Jupiter, he found the eclipses occurring 16⅔ minutes too soon. The only explanation that he could see for the discrepancy was that light required 16⅔ minutes to travel the additional distance it had to go when the planets were on opposite sides of the sun, a distance which was equal to the diameter of the earth's orbit. From these data, the speed of light is determined as follows:

$$Speed = \frac{Distance\ traveled}{Time\ to\ travel\ that\ distance} = \frac{2 \times 93{,}000{,}000}{16\frac{2}{3} \times 60}$$

$$= \frac{186{,}000{,}000}{1{,}000} = 186{,}000\ miles\ per\ second.$$

Römer, not having accurate information on the distance of the earth from the sun, got a figure of 192,000 miles per second.

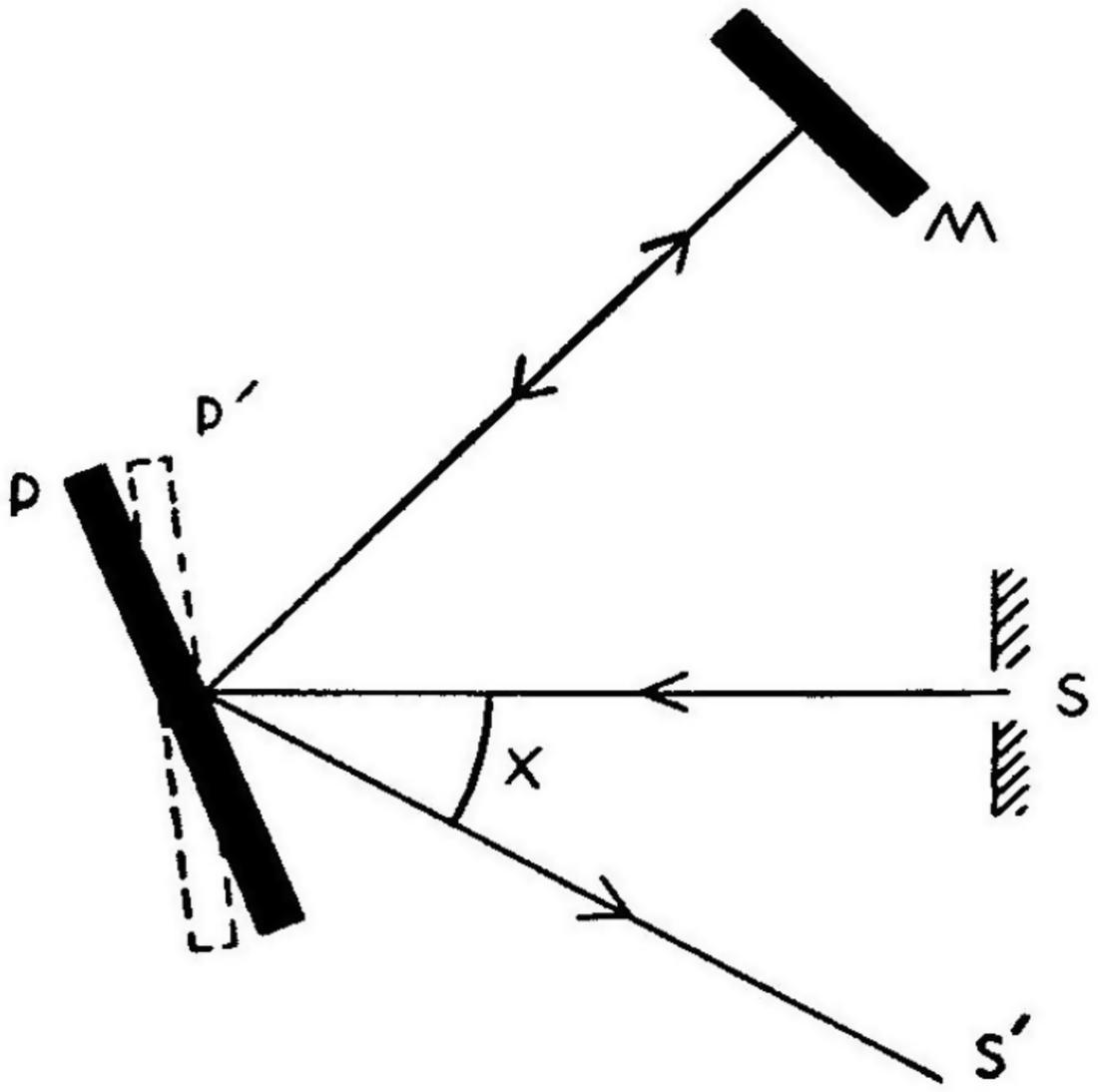


FIG. 189. Simplified diagram of apparatus used by Foucault in determining the speed of light.

The best modern determinations of the speed of light have been made using methods that are modifications of a technique first introduced by Leon Foucault (1819–1868). Foucault was the French physicist who also is noted for having invented the gyroscope, and for having demonstrated that a pendulum persists in its direction of swing regardless of the motion of the earth. Fig. 189 is a bird's-eye view of the essential parts of the Foucault device for measuring the speed of light. *P* is a mirror which can be made to rotate about an axis perpendicular to the paper. *M* represents a fixed mirror. Light from source *S* is reflected from *P* up

to *M*, which it strikes perpendicularly. If *P* is stationary the light will be reflected right back along the same path to *S* again. But let us suppose mirror *P* to be rotating rapidly at a known speed. Now, during the time it takes the light to go from *P* to *M* and back, *P* will have turned through a small angle and the light will be reflected down to *S'* instead of to *S*. By measuring the angle marked *X*, one can determine the small angle through which the mirror turned. The value of this angle and the mirror's rate of rotation are sufficient data for calculating the time it took the mirror to turn through the angle. Since this is also the time required for light to travel from *P* to *M* to *P*, dividing this distance by the time interval gives the speed of the light. In recent years the method has been modified in various ways, the most important work being done by the late Prof. Albert Michelson (1852–1931). His principal improvements were the use of an eight-sided mirror and a considerable increase in the distance between the two mirrors. In a determination made in 1926 the mirrors were located one on Mt. Wilson and the other a little more than 22 miles away on Mt. San Antonio, giving a total light path of some 45 miles. At the time of his death in 1931, Michelson was at work on another measurement of the speed of light using an evacuated tube a mile long. Mirrors were so arranged at the ends of the tube that the light traversed its total length some ten times. The determination was completed three years later by Pease and Pearson who had been associated with Michelson in the experiment.

The value generally accepted today as being the most accurate for the speed of light is,

186,310 miles per second, or
299,776 kilometers per second.

For the purposes of this book it will be sufficient for the reader to remember 186,000 miles per second and 3×10^{10} centimeters per second.

Straight Line Propagation and Diffraction

We commonly think of light as traveling in straight lines, and the fact that it casts sharp shadows seems to demonstrate the correctness of this view. Another illustration is the manner in which light from an object passing through a small hole will produce an image on

a screen. The so-called pinhole camera is a device which can be use in demonstrating this experimental fact. Instead of a lens, there i simply a small hole in the side of the camera opposite the film. Th photograph shown in Fig. 190 was taken with such a device. Th sharpness of the image indicates that the light rays must have travele in fairly straight lines from the object to the film. It is in this con



FIG. 190. Photograph taken with a pin-hole camera. (Photo by F. F. Householder.)

nection that we have what is apparently one of the most obvious differences between light and sound. When light streams out of the doorway of a house there are sharp shadows which define definite limits to the illuminated area. But if radio music is coming through the same doorway there are no " sound shadows " to differentiate sharply a region where sound is audible from one where it cannot be heard. This observed difference was one of the principal reasons why Newton and his contemporaries accepted the idea of light consisting of particles.

It will be noted that in the above lines we referred to this difference between light and sound as " apparently " existing. As a matter of fact, careful investigation shows that light does bend around corners to a small extent. Such bending of a light wave as it passes the edge of an obstacle is called diffraction. Its existence may be demon-

strated in the manner pictured in Fig. 191. Light from a small source, a pinhole in this case, is used to cast the shadow of an object on a screen. Under such conditions the shadow, instead of having sharp edges, exhibits fringes or so-called "diffraction bands." The photographs of a group of small objects, taken in this manner and exhibiting these bands, are shown in Fig. 192. Thus light, like sound, does go around corners. The difference between the two is one of amount rather than

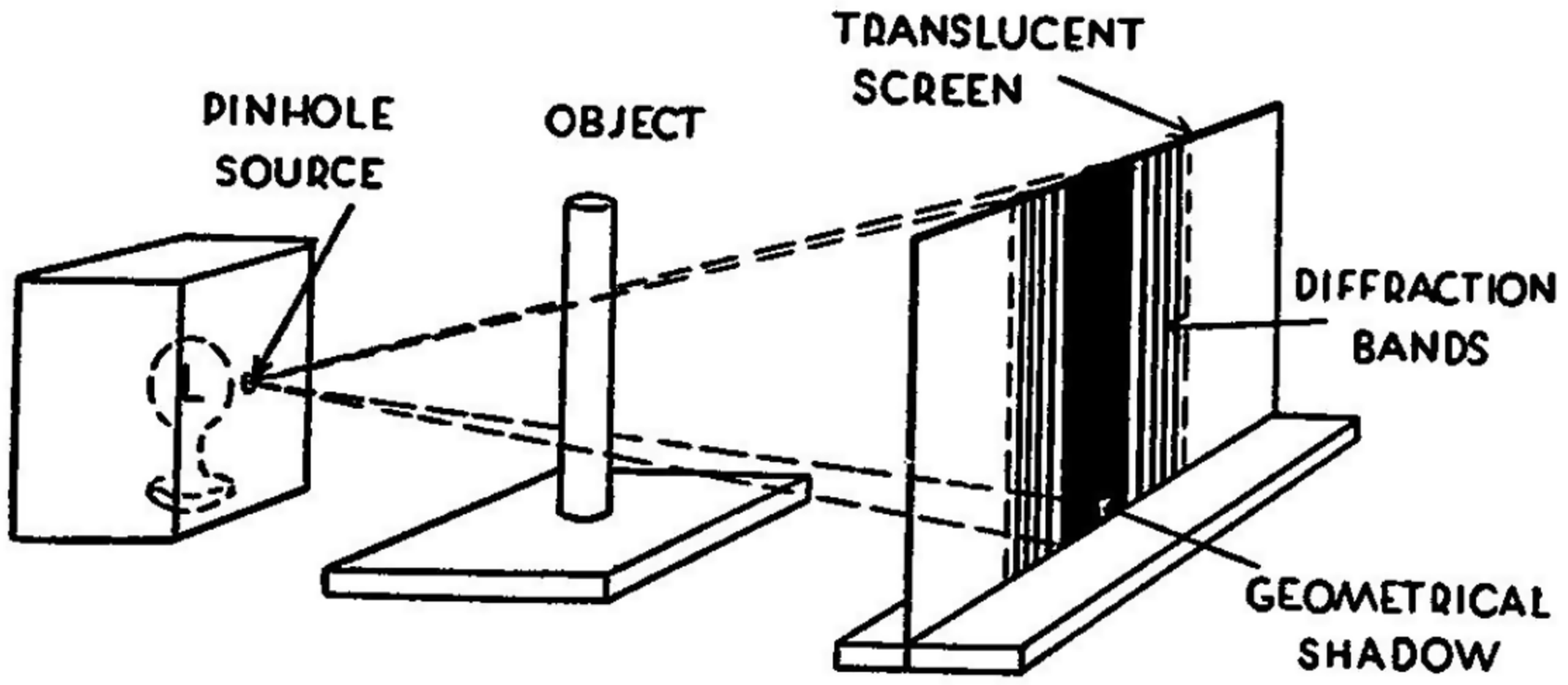


FIG. 191. Diagram illustrating the diffraction of light. (After White.)

of kind. Experiments show that in any kind of wave motion, the shorter the wave, the sharper the shadow it casts. This is true in sound where it can be shown experimentally that tones of high pitch are cut off by obstructions much more sharply than are low ones. In light, where wave lengths are extremely short, the bending is still less and so can easily be missed.

This type of bending of waves must not be confused with *refraction*. Diffraction occurs when waves pass the edge of an obstacle, refraction when they go obliquely into a medium in which they travel at a different velocity. The latter phenomenon has already been described as it occurs in sound and will be considered shortly in the field of light.

Reflection

Reflection is probably the most familiar of all phenomena in this field, since except for actual light sources, almost every object we see is visible because of its ability to reflect light. The relation of this fact to the color of objects will be considered in the next chapter. As has already been suggested, the law of reflection that was stated in the preceding chapter applies to light as well as to sound. This law is illus-

trated in Fig. 193 which shows an actual photograph of light being reflected from a plane mirror.

It is important to recognize that the law of reflection holds for all cases of the phenomenon whether it be from polished reflectors or from so-called diffuse reflectors. The latter term includes all objects that reflect light but do not give mirrorlike images. Some of these have fairly high reflection factors as, for instance, white blotting

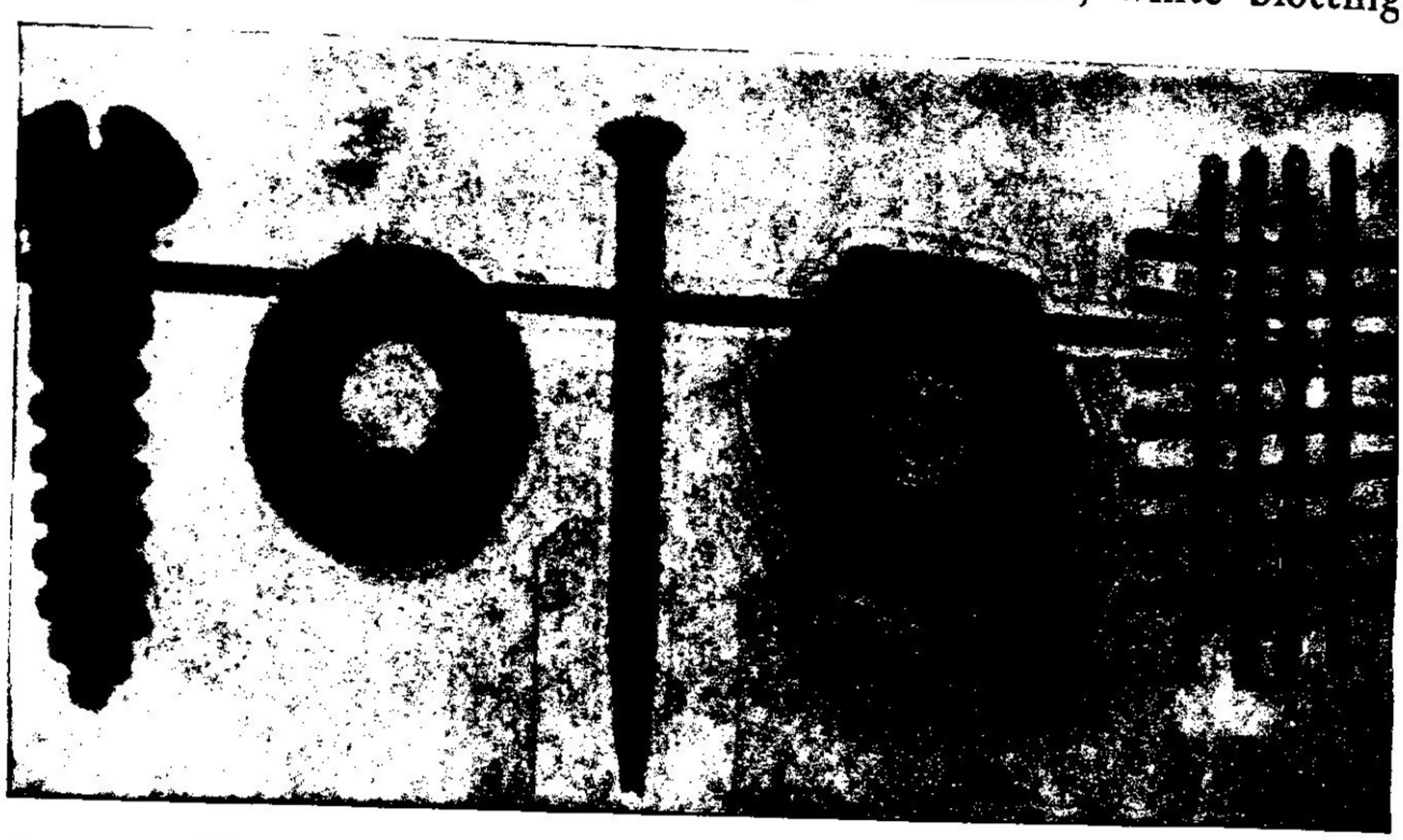


FIG. 192. Photographs of shadows cast by small objects. Note diffraction bands around edges. (From "Classical and Modern Physics," by Harvey E. White, D. Van Nostrand Co., 1940.)

paper, which reflects some 60 per cent of the light that falls on it, and snow, with a reflection factor of 70 per cent to 80 per cent. The surfaces of diffuse reflectors are sufficiently irregular so that parallel rays of light are reflected haphazardly in all sorts of directions and so do not form a sharp image.

The applications which man makes of polished reflectors are numerous and include all uses to which mirrors are put. An interesting point in connection with full-length mirrors is illustrated in Fig. 194. Let us consider first, the man located at position (*a*). *M* is a vertical mirror in which he is looking at himself. The dotted lines represent lines of vision and are drawn in accordance with the law of reflection. To see his eyes and the top of his head he uses a point in the mirror approximately his own height from the floor. To see his feet he uses, as indicated, a point on the mirror halfway between the floor and his

eyes. The location of his reflected image is shown at (a'). This means that a full-length mirror for an economical 6-foot man needs to be only 3 feet long if it is hung with its upper edge about 6 feet from the floor. Now suppose we move this man back to position (b). Following through the same procedure of drawing lines of vision we find

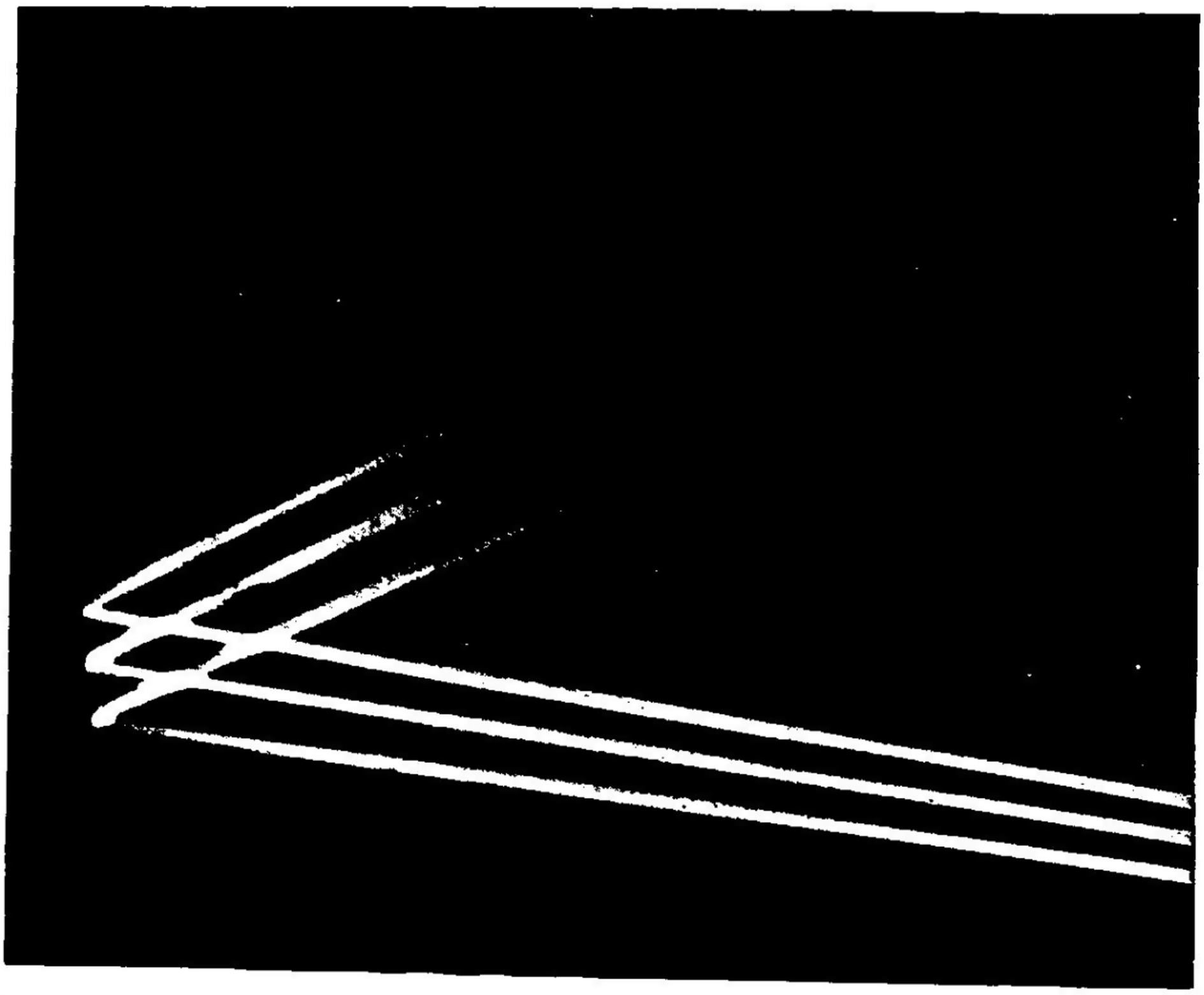


FIG. 193. Photograph of light rays reflected from a plane mirror. (From *Optics and Wheels*, General Motors Corp.)

that the same mirror in the same position will still serve him as a full-length reflector. It is evident then that the distance he stands from the mirror has nothing to do with the problem. In other words, backing away from a vertical mirror does not enable one to see any more of one's self. Since experience has shown that almost no one believes the preceding statement when he first hears it, it would be wise for the reader to find a vertical mirror and test it for himself. Note that what we have said applies only to a mirror that is perpendicular to the floor.

The reflector buttons used so extensively today in highway markers and roadside signs exemplify an application of light reflection which involves the use of three plane mirrors. The mirrors are arranged in

such a manner that each one is perpendicular to the other two. In such a combination a beam of light that enters from any direction will emerge, after reflection, parallel to the entering ray. Regardless of the angle at which an automobile's headlight beams strike the device, the light will be reflected directly back to the driver.

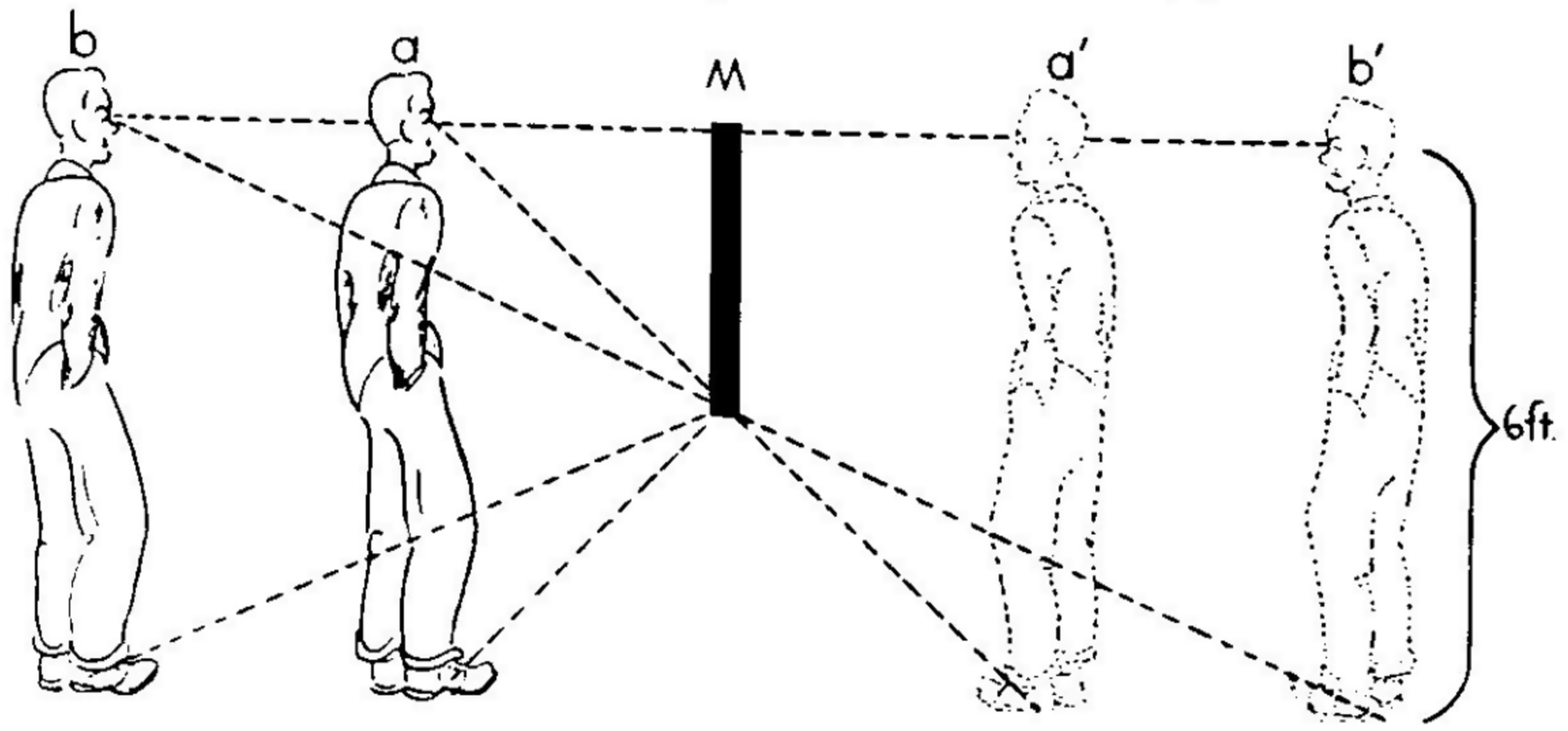


FIG. 194. Diagram showing how a 3-foot mirror can serve as a full-length mirror for a 6-foot man.

One application of the reflection of light from curved mirrors was discussed in Chapter 5. We found there that a concave mirror could be used in a reflecting telescope to bring parallel rays of light to a focus. The photograph shown in Fig. 195 illustrates essentially the reverse of this. Here the source of light is placed at the focus of a parabolic mirror of the sort used as a reflector in an automobile head-

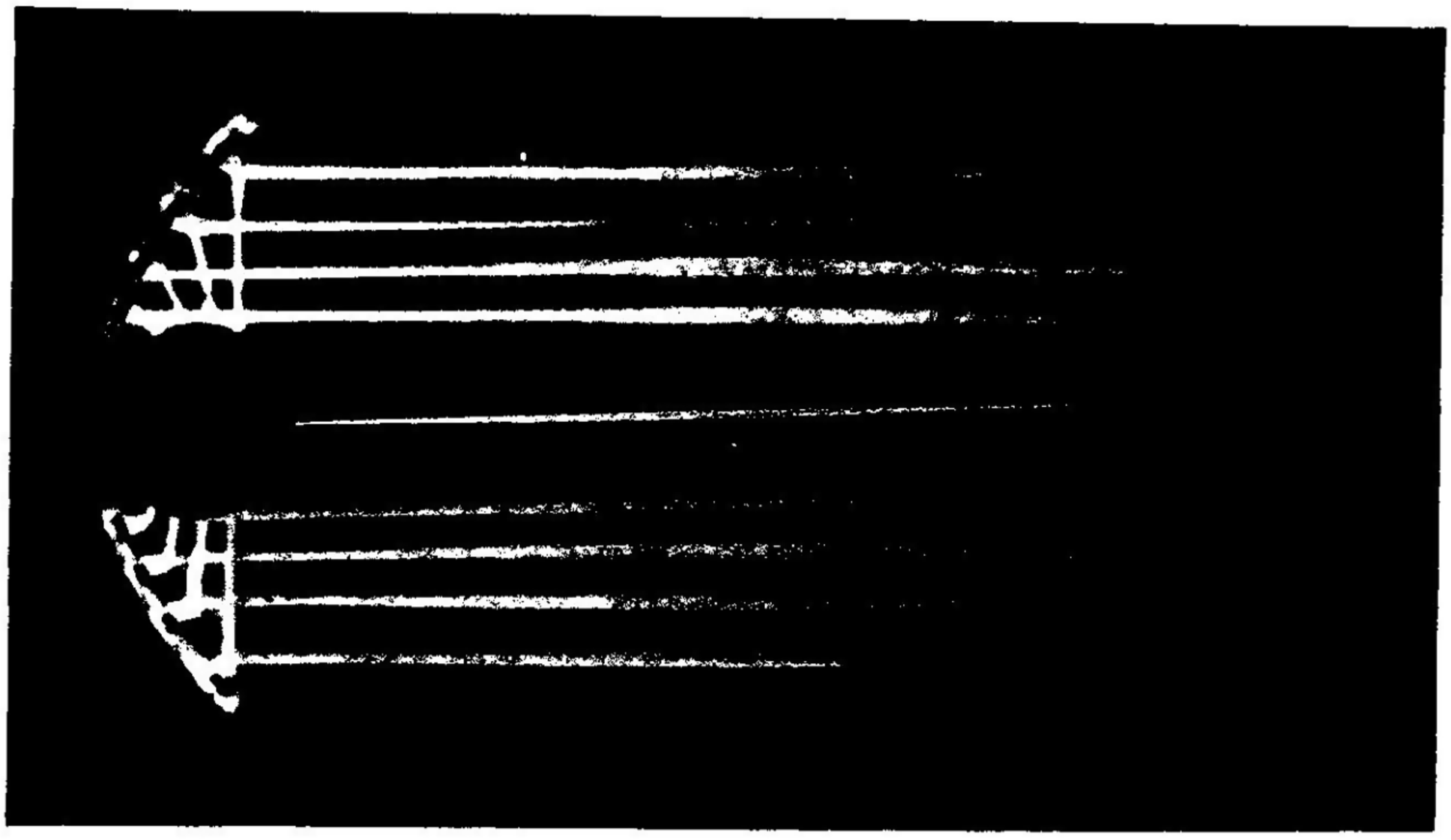


FIG. 195. Reflection by a parabolic mirror of light rays coming from a source located at the focus of the mirror. (From *Optics and Wheels*, General Motors Corp.)

light. The reflected rays now are seen to come out parallel. Note that the law of reflection is obeyed at the points where the rays strike the mirror's surface.

Highly polished metal surfaces make the best reflectors. Silver and aluminum lead the list, with reflection factors of approximately 96 per cent. The best silvered plate-glass mirrors reflect something like 90 per cent of the light that falls on them.

Refraction

When light strikes the surface of a transparent material like glass or water, part of it is reflected according to the law just discussed, and

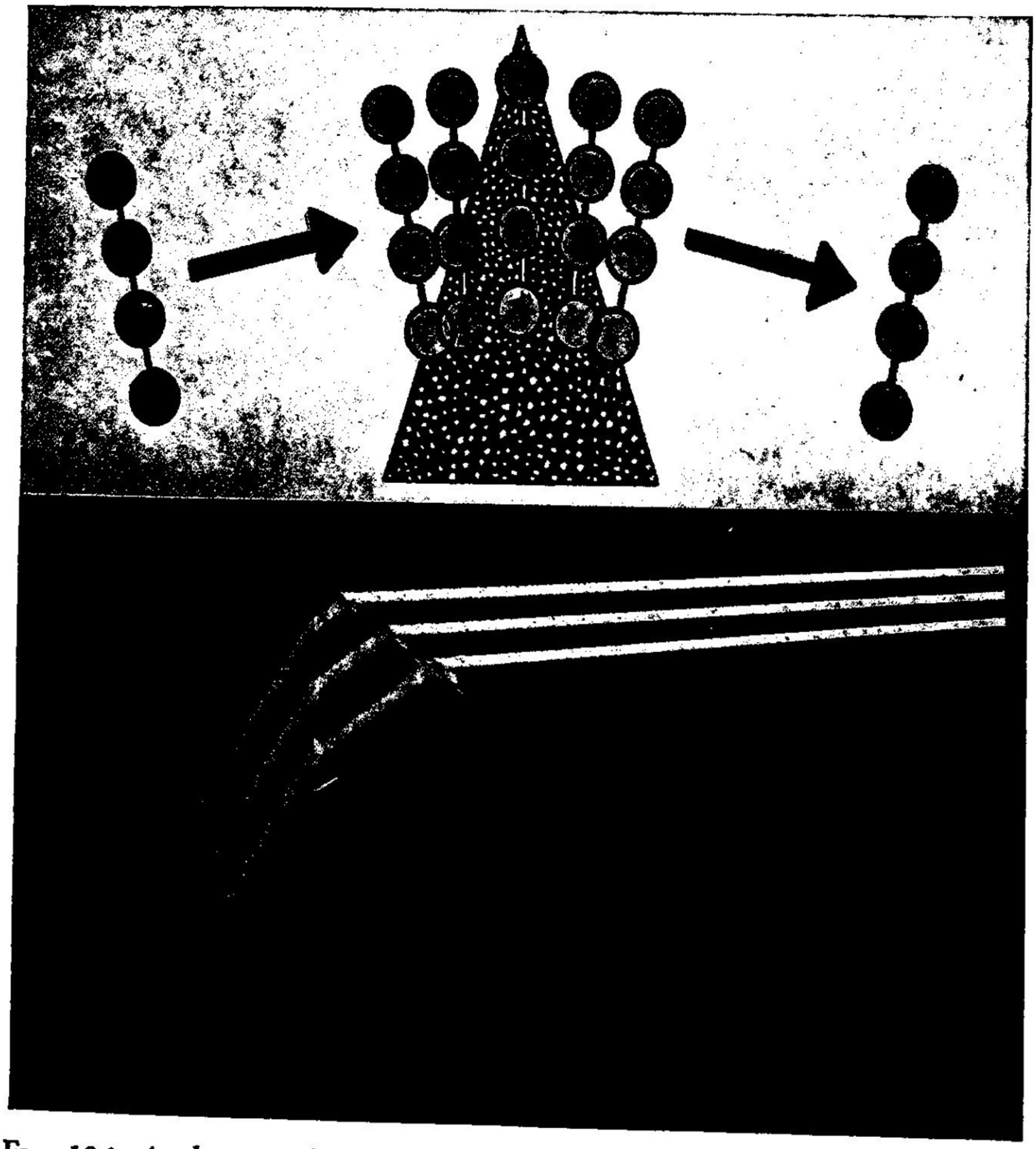


FIG. 196. A photograph of light refracted by a prism and a diagram illustrating the reason for the refraction. (From *Optics and Wheels*, General Motors Corp.)

part of it is refracted into the new medium. It will be recalled that refraction was defined, in the preceding chapter, as the change in the direction of propagation of a wave that occurs when it passes obliquely from one medium to another in which it travels at a different velocity. It happens that light goes through either glass or water more slowly

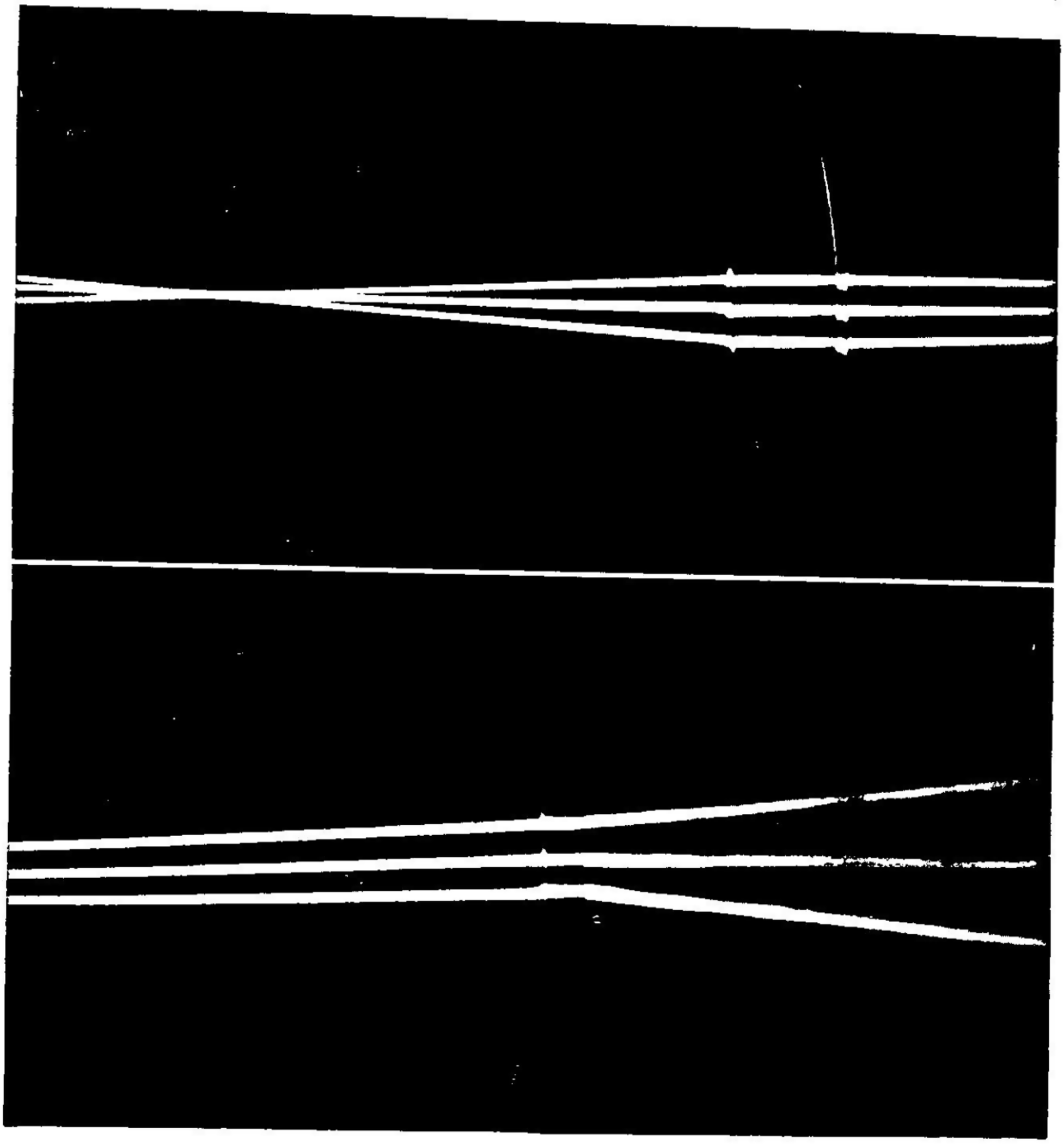


FIG. 197. Refraction of light by (a) a double convex or converging lens, and (b) a double concave or diverging lens. (From *Optics and Wheels*, General Motors Corp.)

than it goes through air, and so when it passes obliquely from air into one of these it is bent exactly as sound waves are bent under similar conditions. If the reader does not fully recall the illustration of the marching soldiers used in the preceding chapter to explain the direction of the bending he should look back at the first part of the discussion of refraction of sound on pages 555–557. Fig. 196 shows

refraction of light rays by a glass prism. The drawing above the photograph may be thought of as representing soldiers marching into a triangular sandy field. The lower end of the column gets into the field first and is slowed down, causing the entire group to swing around toward the right. Then the upper end gets out of the sand first and, speeding up, increases the amount of the turn. Once the column is entirely free of the sand it proceeds in a straight line. The reader should be able to apply this same line of argument to explain the bending of light by the convex and concave lenses pictured in Fig. 197. Note that the former brings the rays to a focus while the latter causes them to diverge. It is the double convex type that is used in the refracting telescope discussed in Chapter 5.

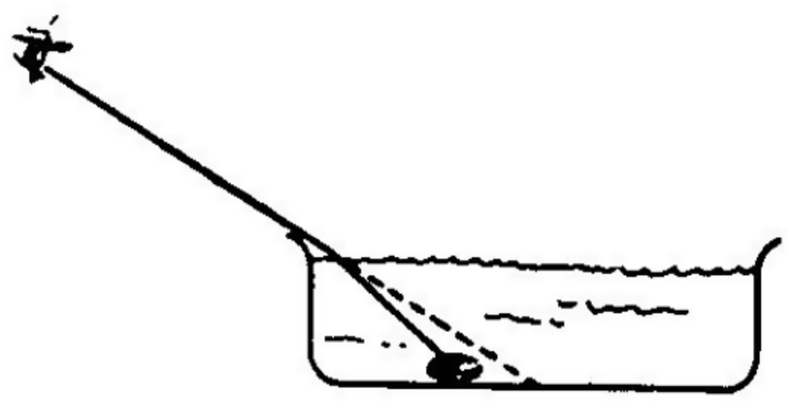


FIG. 198. Demonstration of refraction. Coin cannot be seen when vessel is empty but becomes visible after water is poured into it.

An easy way to demonstrate the fact of refraction of light for one's self is pictured in Fig. 198. A coin is placed on the bottom of a shallow basin at a place where the sides of the container just hide it from view. When the vessel is filled with water the coin becomes visible, although neither it nor the observer has moved. The reason lies in the refraction of the light rays coming from the coin because of the change in medium. There are a number of commonly observed refraction phenomena. The twinkling of stars results from refraction of the starlight as it passes through regions of the atmosphere which are at varying temperatures. One of the most familiar examples of refraction is the appearance of a straight piece of wood, such as an oar, that is partly in water. Everyone has noticed the apparent bend or break at the surface of the water in such a case.

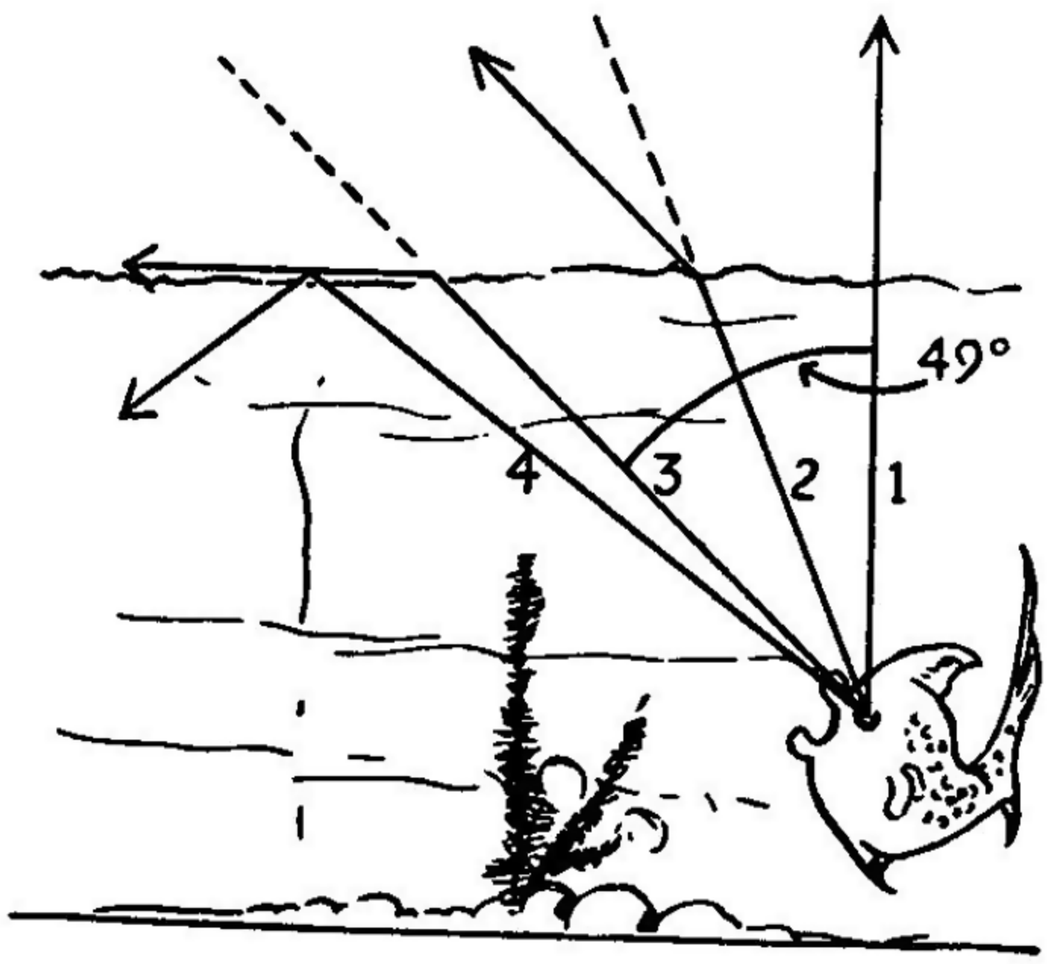


FIG. 199. Diagram illustrating phenomenon of total reflection and meaning of critical angle.

One of the most interesting of refraction phenomena goes by the name of " total reflection." It is illustrated in Fig. 199 which shows a fish looking out of the water in various directions. Note that it is looking from water, a " slower " medium, into air, a " faster " medium. When it looks in direction *1*, there is no bending at the surface of the water because the ray strikes perpendicularly. In direction *2*, there is some refraction and the fish sees objects in the direction indicated by the solid line. The amount of this refraction increases with the angle until the point shown by direction *3* is reached. Here, although the fish's angle of vision is only 49 degrees from the vertical, it sees the horizon. This angle is called the *critical angle* and varies with the transparent substance. In the case of common glass, for instance, it is 42 degrees. Thus the fish sees everything from one horizon to the other within an angle of 2 × 49, or 98 degrees, instead of within 180 degrees as is the case with us. But suppose our fish directs his gaze at an angle greater than 49 degrees. In that case, he is unable to see out of the water at all, but gets instead a reflected view of something in the bottom of the pond in the direction shown by ray *4*. So the phenomenon of total reflection occurs when light, going from a slower into a faster medium, strikes the boundary at an angle greater than the critical angle. One manner in which light rays are reflected totally in a glass prism is illustrated in Fig. 200. An application of this effect is on page 610. It should be emphasized that there is no silvering whatever on the surfaces at which the reflection occurs. It is purely a refraction phenomenon occurring as described above.

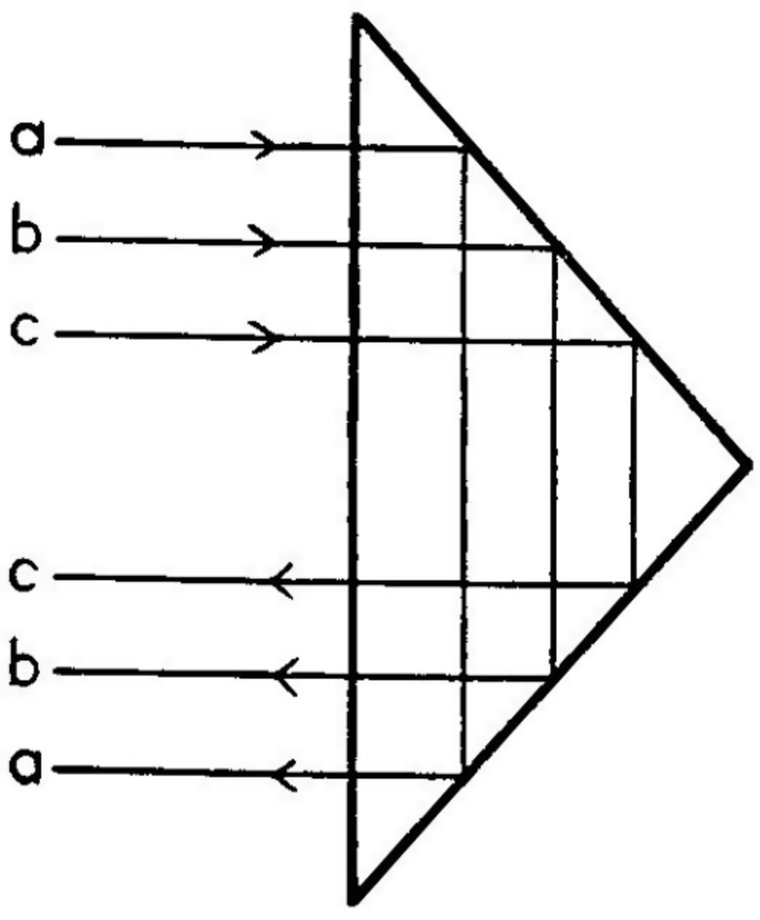


FIG. 200. Total reflection by a prism. (After White.)

It should be noted that our use of the word medium in the above discussion is not entirely justifiable. The term as we used it in sound referred to the actual material in which the waves were propagated. This obviously is not the case with light in air, glass, or water, since they can all be removed without hindering transmission of light. However, it is true that light goes through a vacuum a little faster

than through air, and more rapidly through air than through either glass or water.

Interference

At first thought it might seem unlikely that two beams of light could be combined in such a way as to produce darkness. Nevertheless, we have already seen in the phenomenon of beats, that two sound waves can be made to add together to give moments of silence which alternate with intervals of unusual loudness. And it will be recalled

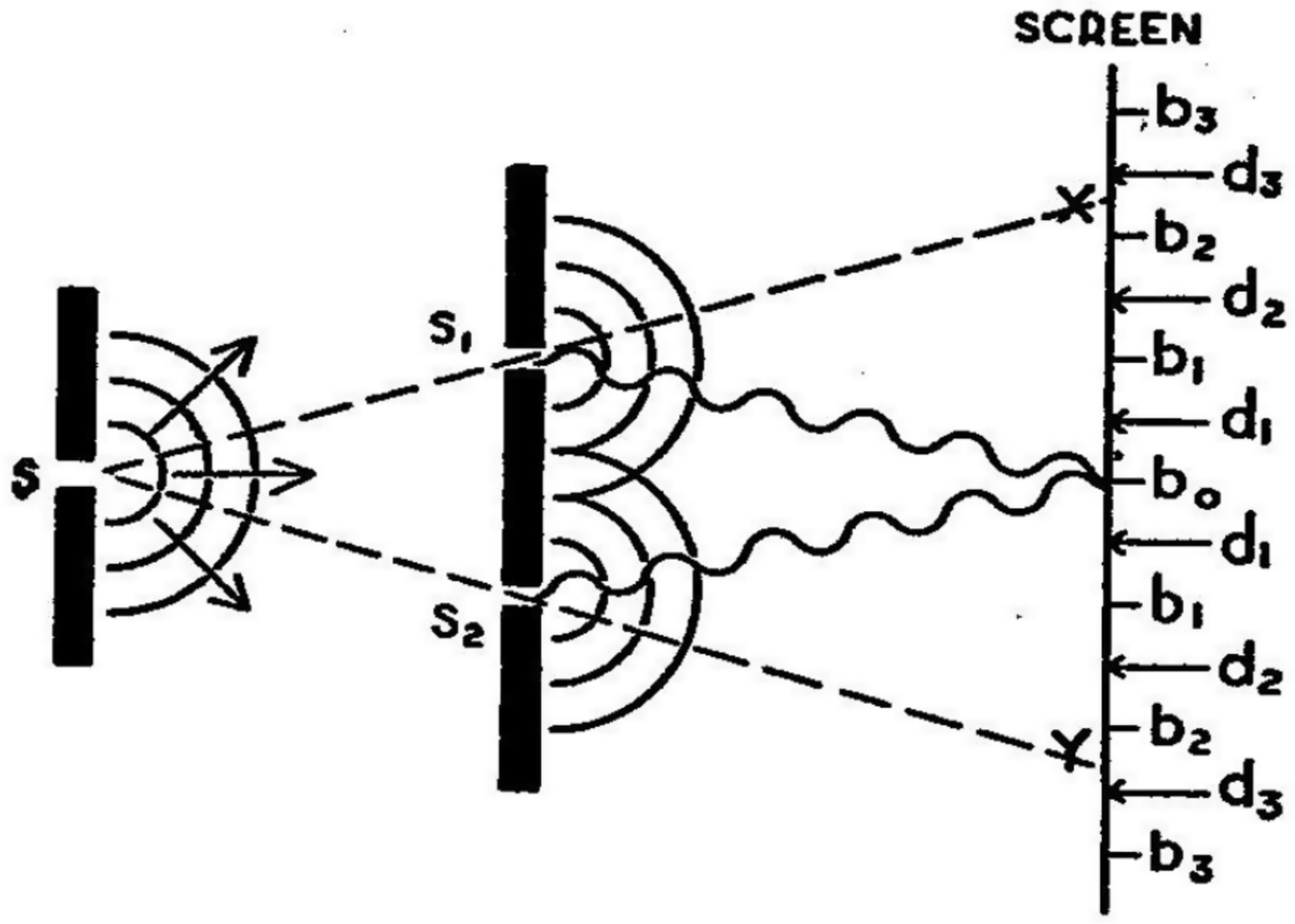


FIG. 201. Diagram illustrating the double-slit experiment for demonstrating interference of light.

that loudness in sound corresponds to brightness in light. The experiment which first demonstrated the existence in light of a phenomenon analogous to beats was performed in 1801 by an English scientist named Thomas Young (1773–1829). As we shall see a little later, the results of this experiment seemed at that time to settle forever, in favor of the wave theory, the question of the nature of light. Fig. 201 shows schematically the set-up for Young's so-called "double-slit" experiment. We shall describe the experimental results first, and then ask how they can be explained on the basis of light being a wave motion. In the diagram, S represents a light located back of a narrow slit, the purpose of which is to provide a small source. S_1 and S_2 are other slits equidistant above and below the level of S. Now we may ask what sort of pattern of light should appear on the screen at the right. It might seem reasonable to expect to find bright spots only at points X and Y which are in line with the slits and the original source. Actually, when

the experiment is performed, one finds a whole series of alternating bright and dark bands with the brightest one appearing at b_o. From this point out in either direction they grow increasingly faint. Fig. 202 is a photograph of an interference pattern resulting from such an experiment. Let us go back to the apparatus diagram and see whether these results make sense if light consists of waves. It is true in any sort of wave motion that if the waves go through a small opening in a screen, that opening acts exactly as though it were a new source. In other

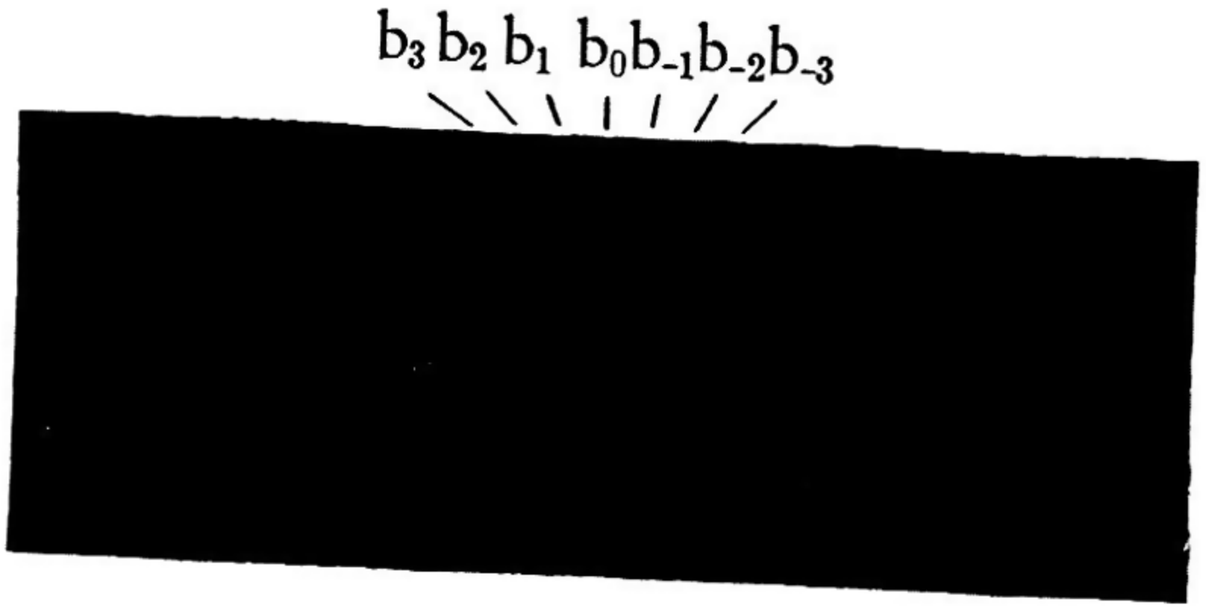


FIG. 202. Photograph of interference fringes produced by a double-slit experiment like that shown in Fig. 201. (From "Classical and Modern Physics," by Harvey E. White, D. Van Nostrand Company, 1940.)

words, light waves would spread out in all directions from S. Similarly, after light passed through S_1 and S_2, they would in turn act like new sources of light waves. Now let us fix our attention on the two waves drawn from S_1 and S_2 to b_o. These would start out in phase, or in step, from the two slits since S_1 and S_2 are equidistant from S. Also, since the distance from S_1 to b_o equals that from S_2 to b_o, they would arrive at the screen in phase; that is, crests or troughs from the two waves would come together at b_o. This, as we saw in the case of sound, would produce reinforcement which in light means unusual brightness. However, a little above b_o there would be a point on the screen where the distance to S_2 would be just one-half a wave length longer than that to S_1. At this point, indicated by d_1, waves would arrive exactly out of phase; that is, a crest from one would join a trough from the other. Such a situation in sound produces silence; in light it produces darkness. Still a little higher on the screen is b_1, which is exactly one wave length farther from S_2 than S_1, giving reinforcement again, and so on. Obviously, there would be a similar series of points below b_0. Thus, the experimental results are just what one should expect if light consists of waves.

This interference effect can be seen rather easily by scratching two narrow lines in the emulsion on a photographic plate or film and then observing a vertical filament in a clear light bulb through this double slit. If the light source is white, the fringes will appear colored. If a red or violet filter is placed in front of the source, red or violet fringes will be seen respectively, with the red ones appearing farther apart than the violet. This variation in separation means that the different colors have different wave lengths. From the dimensions of the apparatus used in Young's experiment and the distance between the interference fringes it is possible to calculate the wave length of the light coming from S. Such determinations give the following values for the lengths of the waves at opposite ends of the spectrum:

Wave length of red = .000,07 *cm.* = .000,028 *inches*
Wave length of violet = .000,035 *cm.* = .000,014 *inches*

The wave length of green light has already been given back in Table 28 as .000,02 inches.

The Doppler Effect

The Doppler effect, as it occurs in light, consists of an apparent small change in frequency or color because of the motion of the source or of the observer. The shift is toward the violet end of the spectrum if the source of light is approaching, or toward the red end if it is receding. The reason for the phenomenon should be clear to the reader from the discussion of it in the preceding chapter on sound. The application of the Doppler effect to the measurement of star velocities was discussed in Chapter 5. It should be emphasized that the Doppler shift in light must be detected by careful wave-length measurements. It is not sufficiently large so that one would notice any change in color with the unaided eye.

Polarization

The phenomenon of polarization has received a considerable amount of popular attention in recent years because of a number of important and interesting practical applications that have been made of it. From the standpoint of the nature of light it is worth some consideration, because it provides the principal evidence for the idea that

light waves are transverse in nature. Polarization can be described most easily in terms of a rope-wave analogy. In Fig. 203 two cases are shown of a rope threaded through slots in a pair of boards. Let us consider part (*a*) first. Somewhere to the left of the diagram there is assumed to be a vibrating source that is sending transverse waves along the rope from left to right. Also it is to be assumed that this

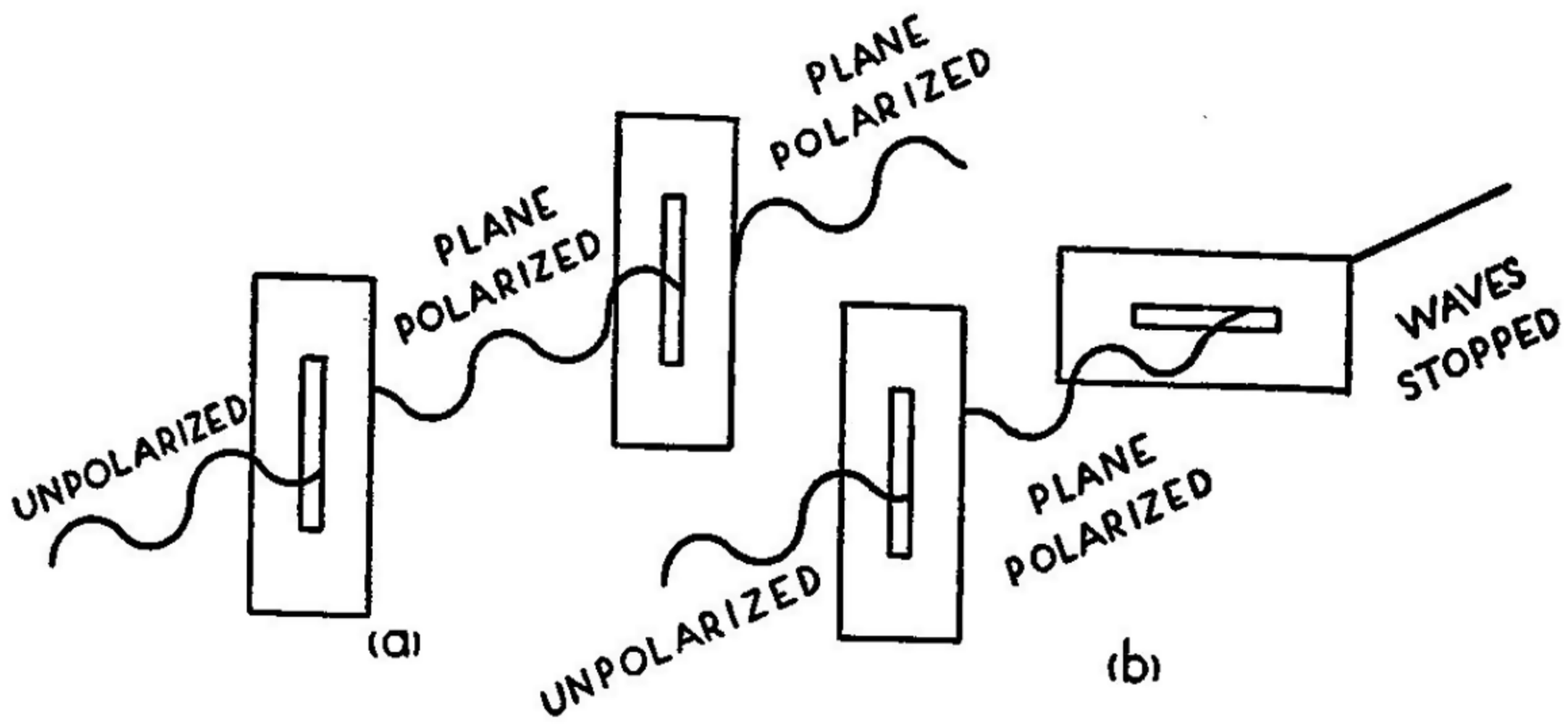


FIG. 203. Diagram illustrating polarization of rope waves.

source vibrates in various planes, that is, sometimes it moves up and down in the plane of the paper, sometimes back and forth at right angles to the paper and sometimes in directions in between these two. Therefore, the particles of rope to the left of the first board also are moving in a variety of directions, all at right angles to the direction of propagation of the waves. Such waves are said to be *unpolarized*. It is obvious that, of all these waves, the only ones that can get through the slot in the first board are the ones for which the vibrations are up and down. Therefore, between the boards, all the vibrations will be in a vertical plane, that is, the plane of the paper. Such waves are said to be *plane polarized*, and the device which produces the effect, the slot in the board in this case, is called the *polarizer*. Since the second slot is parallel to the first, the polarized waves will go unhindered through the right-hand board. Now we shall turn to part (*b*). It differs from the first part in that the second board has now been rotated through an angle of 90 degrees making the slots perpendicular to each other. Again, the first slot polarizes in a vertical plane the waves coming from the vibrating source. It is evident that these will be unable to go through the horizontal slot in the second board, so that beyond

it there will be no waves at all. It should be noted at this point that if the waves had been of the compressional type, as for example the ping-pong ball arrangement described in Chapter 28, rotating the second slot should have had no effect on their passage.

There are materials which appear to function with respect to light exactly as the above slots do for rope waves. The best known of such substances today is Polaroid glass. Fig. 204 shows an arrangement of two Polaroid disks and a flashlight source analogous to that pictured for rope and slots in part (*b*) of the preceding figure. In other words, the disks are "crossed" and almost no light gets through the second one. If, starting with this arrangement, one rotates the second disk, light begins to come through it, growing in intensity until it reaches a maximum after a rotation of 90 degrees. The second disk is called the analyzer, since with it one can tell whether or not the light is polarized. The two disks are alike and could be interchanged just as could the boards in the analogy. Each disk by itself is transparent. This fact is illustrated in Fig. 205 which shows two of them in the "crossed" position overlapping a little. Thus, one can say that *if* light consists of transverse waves, and *if* these polarizing materials behave toward it in a fashion analogous to that described for slots and rope waves, *then* the phenomenon of polarization in light becomes understandable. If the reader has properly appreciated the implications of the scientific method, he will understand that this is not quite the same as saying that the demonstration with rope waves *proves* that light consists of transverse waves. To show that a certain mechanism will produce a given result does not prove that it is the only one that will do so. The technique of trying to prove things by illustration or analogy is a fallacious method of argument although, as the reader will recognize, public speakers in many fields make extensive use of it.

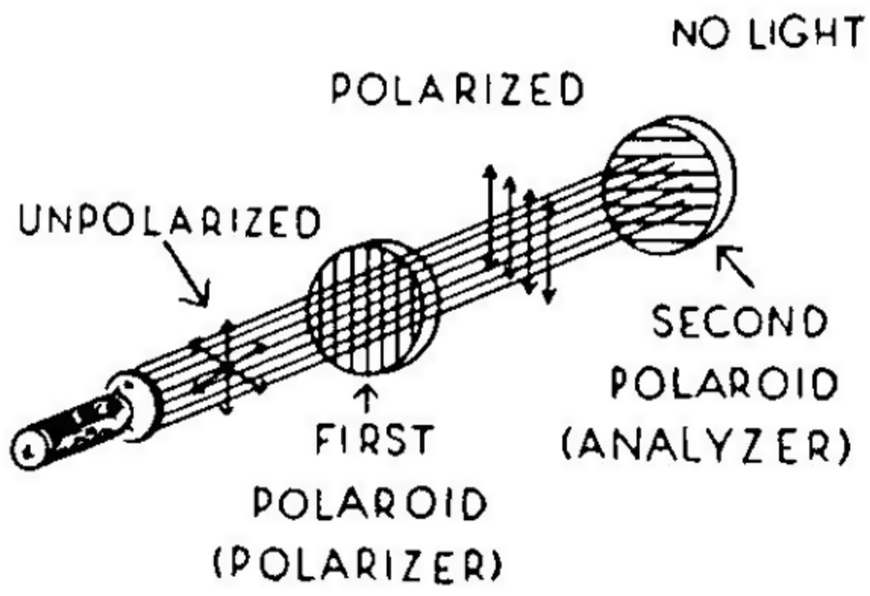


FIG. 204. Diagram illustrating polarization of light. Polaroid disks are arranged analogous to (b) of Fig. 203.

There are several methods by which light can be polarized, other than by the use of Polaroid disks. When unpolarized light strikes a

polished glass surface at an angle of about 57 degrees, the reflected beam is almost entirely plane polarized. With other angles of incidence the reflected light is partially polarized. Crystals of such materials as calcite, quartz, mica, and ice have the property of splitting a beam of light into two beams, both of which are plane polarized but

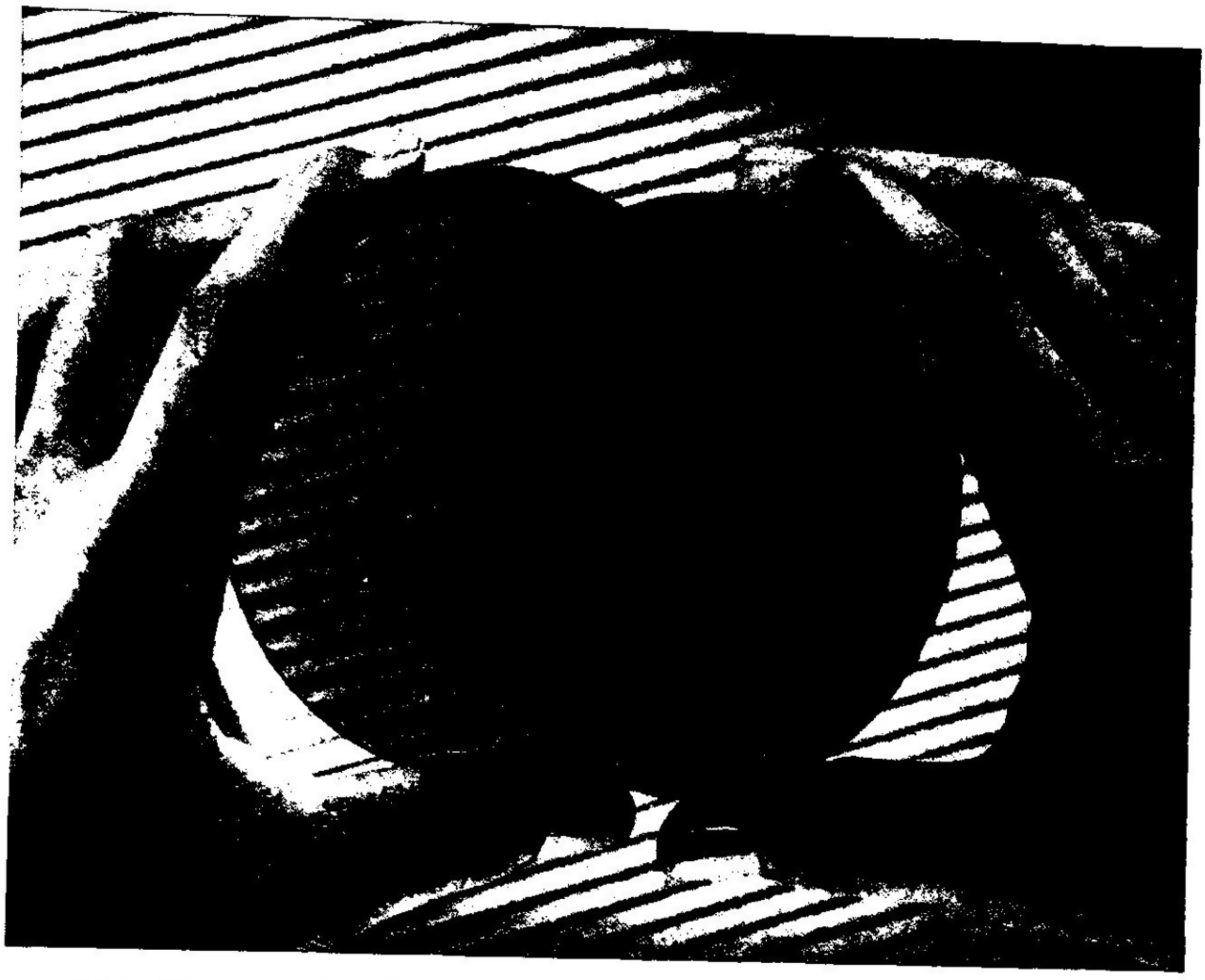


FIG. 205. Photograph of light coming through "crossed" Polaroid disks which overlap. (Polaroid Corp.)

at right angles to each other. Other methods of polarizing light include passing it through tourmaline, and through a solution containing particles which scatter the light. The dust in the atmosphere, for instance, partially polarizes the light that is scattered by it.

Many of the more recent applications of polarized light have been made possible by the development of Polaroid materials, which for the first time has permitted the manufacture of sheets of polarizing material larger than an inch or two in diameter. Uses which involve the reduction of glare are probably more or less familiar to the reader. The light from a reading lamp that has polaroid glass surrounding the source produces less glare than does unpolarized light. Polarizing sun glasses act in a similar fashion. Glare from automobile headlights can be es-

sentially eliminated by using polarizing headlight lenses, and then making the windshield, or a section of it, of a polarizing material that is " crossed " with respect to the lenses of approaching headlights.

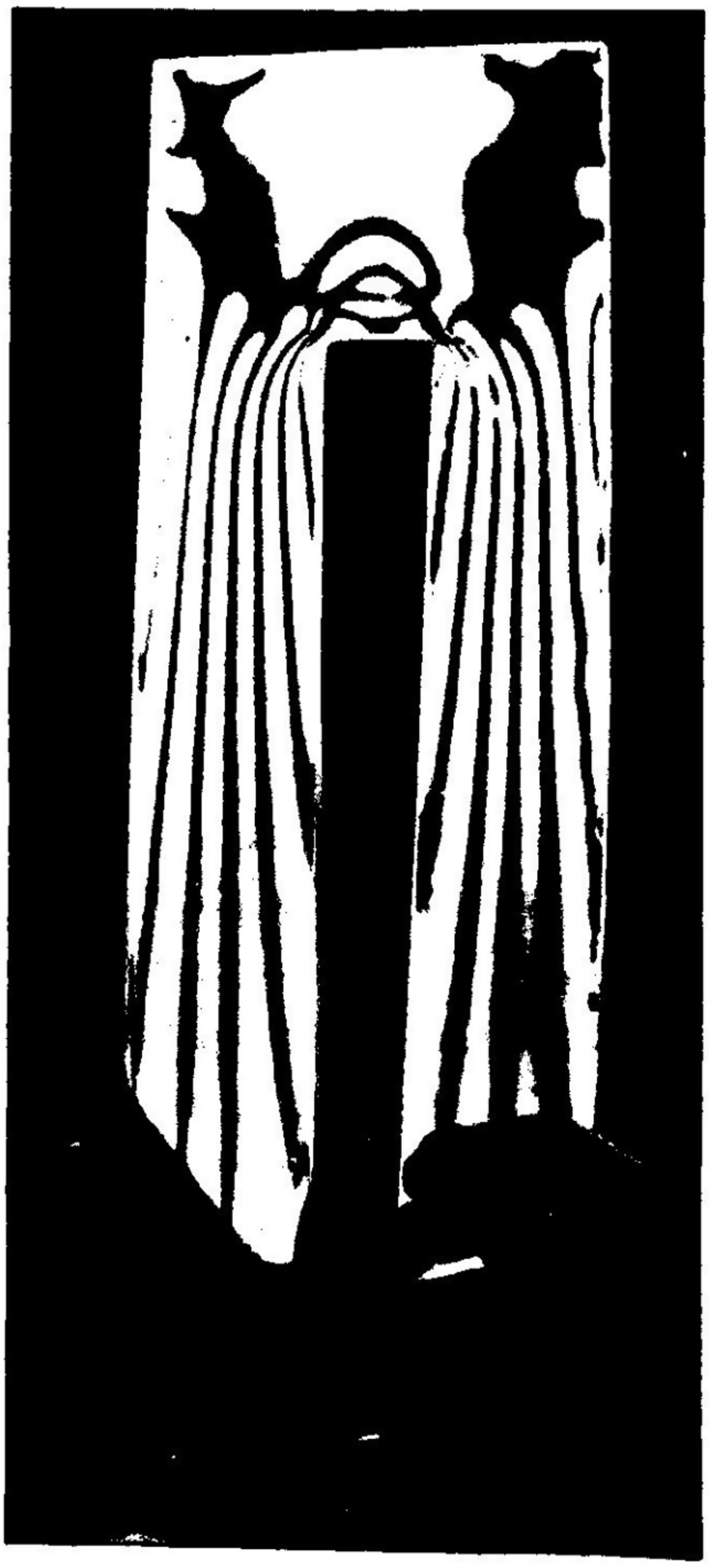


FIG. 206. Photograph of a transparent U-shaped object illuminated with polarized light while under strain. (Polaroid Corp.)

Important engineering applications of polarized light are based on the experimental fact that transparent materials under strain will exhibit vivid color effects when illuminated with polarized light. These color markings are such as to make it possible to locate points of maximum strain in the substance. This phenomenon is illustrated in Fig. 206 which shows a U-shaped transparent object with force applied to the ends while the piece is illuminated with polarized light.

The color effects are so distributed as to make it possible to locate points of maximum strain. In this way, if one wishes to know the places of strain in a bridge member or a gear, one can make a scale model of the piece out of transparent material, apply loads to it proportional to those that the completed member will experience, and study the piece under polarized light.

THE NATURE OF LIGHT

Man's curiosity concerning the nature of light was aroused at a fairly early time, just as it was with regard to the nature of heat and of matter. The fact that the sun is the principal source of light made it clear that the transmission of energy from one place to another was

involved in the problem. Now, if the reader will think back over the various examples of energy transmission we have considered, he will see that they all can be reduced to two types, those that involve the actual projection of material bodies and those in which waves go through an intervening medium. For this reason, theories of the nature of light have always pictured it either as consisting of tiny particles or as being a wave motion in some sort of medium. The former are commonly called corpuscular theories and the latter wave theories.

Early Theories

Early Greek scientists and philosophers, with the exception of Aristotle, adopted one variation or another of a corpuscular picture. One group held that the particles were emitted by the eye and that an object became visible when these little bullets struck it. This idea lost favor, however, because by it, one could not explain why people failed to see just as well at night as during the day. Pythagoras and his followers, in the sixth century B.C., reversed things and taught that the particles were shot out by the object which, in this case, was seen when the tiny light bullets struck the observer in the eye. Plato modified this picture about a century later, teaching that sight resulted when a stream of light particles coming from the eye mixed first with a set coming from the sun, and then with another group emitted by the body that was being observed.

Aristotle, in the fourth century B.C., discarded all particle theories and taught that light was a sort of quality or action in a transparent medium through which it was transmitted instantaneously.

Newton's Particle Theory

Sir Isaac Newton (1642–1727) was one of the first to carry out extensive experiments in the field of light. Let us go over the list of phenomena discussed previously and ask what bearing each one, in Newton's opinion, had on the question of the nature of light. Römer's measurement of the speed of light occurred when Newton was around 34 years of age. Since both waves and particles would travel with a finite speed, that particular fact favored neither theory. Reflection was no help either because, as we have seen, the same law applies to both waves and corpuscles. The apparent straight line propagation of

light was perhaps the principal reason for his eventual adoption of a particle theory, although he seems to have experienced a considerable amount of indecision from time to time. Knowing the ability of sound and water waves to bend around corners, Newton was unable to explain the sharp shadows he observed in light on any other than a particle basis. We know today, as has already been pointed out, that in proportion to their size, light waves are diffracted just as much as are sound waves. It must be remembered that the shortest audible sound waves are more than 20,000 times as long as the longest visible light waves.

By assuming that light traveled faster through denser materials like water and glass than through air, Newton was able to account entirely satisfactorily for refraction phenomena. The assumption could not be tested during Newton's lifetime, however, because no one was able to make velocity measurements in media other than air. Interference phenomena had not yet been discovered and not much was known about polarization. One difficulty in the way of Newton's particle theory was the fact that when light strikes something like glass, part of it goes through and part is reflected. For instance, it would be hard to imagine a surface against which one could throw a bucket of ping-pong balls and have part of them bounce off and the rest go through. To account for this observed fact in light, Newton had to assume that part of the time the particles were in what he called a state or mood of " easy reflection," and the rest of the time in one of " easy transmission." Then what happened to any given particle when it struck a surface could be said to depend on which of these states it was in at the moment.

At the time of Newton's death and for a good many years afterwards, the problem of the nature of light was considered to be pretty well settled, with essentially everyone agreeing on a particle picture.

The Wave Theory

Thomas Young is generally credited with having founded the wave theory of light. His experiments on interference of light, performed near the beginning of the nineteenth century, seemed understandable only in terms of light consisting of waves. The nature of the explanation of this has already been discussed in this chapter. Not

only that, but diffraction and polarization, which by Young's time had been more fully investigated, could be adequately explained by waves. This fitting of diffraction into the wave picture, of course, eliminated the " straight line propagation " argument on the particle side. The experimental facts concerning finite velocity and reflection, we have already seen, agree with either theory. What seemed to be the final test came in 1849 when Foucault measured the speed of light in media other than air and found that the more dense the material, the slower light went through it, completely contradicting Newton's basic assumption with regard to refraction. And so again in 1850 the problem of the nature of light seemed to be finally settled, this time in favor of the wave picture.

Present Status of the Problem

In spite of its success in explaining the experimental phenomena just mentioned, even in 1850 the wave theory had one outstanding weak point. That concerned the medium in which light waves travel. The term " ether " has come into common use as the name of this medium but it must be recognized that ether was invented, not discovered. It was invented rather definitely because of the fact that all other types of waves that had been studied required media, and it was felt that, therefore, there must also be one for light. While much work and research has gone into the study of this problem and some advances have been made, this particular feature of the wave theory still offers a number of unsolved problems.

Additional difficulties for the wave theory of light came into being with the startling series of discoveries in physics that began around 1895 and included X-rays, radioactivity, the electron, and photoelectric effect. Because of the complexity of the theories involved in these phenomena, the reader will have to take on faith most that we say here about their bearing on the nature of light. The principal point to be made is that some of the experimental facts in connection with certain of these discoveries, notably the photoelectric effect, are understandable only in terms of light's being corpuscular in nature. Also, the results of certain experiments made on electrons have removed diffraction and interference from the group of seemingly conclusive wave arguments.

Today we have three kinds of experimental facts about light. One group can be accounted for by either a wave or a particle picture; a second seems to make sense only in terms of a wave theory; while the third group can be explained only if one thinks of light as consisting of particles. Since all of the more familiar manifestations of light come under the first and second of these groups, we commonly use the language of waves to describe light phenomena, while recognizing that a future completely satisfactory theory will have to be in agreement with all facts.

In closing this chapter, let us emphasize the point that is the principal reason for going into as much detail as we have in connection with this story of man's theories of the nature of light. It affords an excellent example of the way in which scientific theories develop with the growth of evidence. Here was a problem that seemed entirely settled at one time, then some years later just as completely settled in exactly an opposite manner, and finally has become definitely unsettled again. The important thing to note is that each time the situation changed, that change occurred because of new experimental evidence, and was made only when such new evidence made it necessary. This illustrates again the point that has been made before in this book, namely, that any scientific theory is satisfactory at a given time if it agrees with all the experimental facts known *at that time*, and that to ask whether it is right or wrong is to propound an interesting but irrelevant question. One of the principal functions of any scientific theory lies in the fact that it stimulates men to perform further experiments in order to test it, and its value in that connection is just as great if the knowledge gained in the experiments contradicts it, as it is if the new information supports it.

Questions

Discussion

1. Describe Galileo's attempt to measure the speed of light and explain why it was unsuccessful.

2. Look up the method of measuring the speed of light used by Fizeau and show how it differed from Foucault's procedure.

3. If the moon should suddenly blow up at midnight tonight, at what time would anyone on the earth see the explosion and when would they hear it?

4. Answer Ques. 2 substituting sun for moon.

5. Calculate the length of the shortest possible full-length mirror that a 6-foot man and his 5-foot wife could both use.

6. Where would the angles of incidence and reflection be in the photograph shown in Fig. 193?

7. Can you give a reason why you naturally back away from a full-length vertical mirror in a clothing store even though doing so does not enable you to see any more of yourself?

8. Find some practical applications of curved mirrors other than those mentioned in the chapter.

9. Draw lines showing how light rays would be reflected from a convex mirror. Why might such a mirror be better than a flat one for use as an automobile rear-view mirror?

10. After reading the explanation of refraction in sound in Chapter 29, explain the refraction of light by the two lenses pictured in Fig. 197.

11. Find some examples of refraction phenomena other than those mentioned in this chapter.

12. In terms of one of the light phenomena discussed in this chapter, explain why it is possible to see the moon after it has gone a little below the horizon.

13. Why wouldn't it be a good idea to try to explain going through a red traffic light by saying that as you approached it the Doppler effect made it appear green? Would a more plausible situation be for a green light to appear red? Explain.

14. Calculate about how many times the wave length of middle C is greater than that of green light.

15. Explain the answer that you would give to Ques. 5 in the Multiple Choice group.

16. Calculate the approximate number of octaves, as that term was defined in Chapter 29, there are between violet light and red light.

17. Explain why the phenomenon of polarization seems to be an argument for transverse waves in light.

18. Calculate the frequency of green light.

Multiple Choice

1. Which one of the following phenomena would be least likely to occur in sound waves? (a) Interference. (b) Doppler effect. (c) Refraction. (d) Polarization. (e) Reflection.

2. At the present time the experimental evidence concerning which of the following (one or more) is understandable only in terms of the wave theory of light? (a) Photoelectric effect. (b) Refraction. (c) Polarization. (d) Reflection. (e) Existence of a finite speed.

3. Which two of the following properties of a light wave multiplied to-

gether should give approximately the figure "186,000"? (a) Period in seconds. (b) Wave length in feet. (c) Frequency in vibrations per second. (d) Wave length in miles. (e) Speed in miles per second.

4. List all of the following that are due primarily to refraction of light. (a) Shift in observed frequency of light due to motion of the source. (b) Focusing of light by the Mt. Wilson 100-inch telescope. (c) Bent appearance of a straight stick that is partially submerged in water. (d) Magnifying effect of a concave mirror. (e) Total reflection.

5. Which one of the following names is out of place, on the basis of material discussed in this chapter? (a) Newton. (b) Galileo. (c) Michelson. (d) Foucault. (e) Römer.

True–False

1. Variations in brightness of light are analogous to variation in pitch of sound.

2. Polarization is primarily a refraction phenomenon.

3. An experiment performed during the last century, which provided a test between the wave and particle theories of light, involved the measurement of the speed of light in media other than air.

4. If a photograph of the spectrum of a certain star showed all the lines shifted a little toward the red from their normal position, one would conclude that the star and the earth were getting closer together.

5. We see light of widely differing wave lengths as light of different colors.

6. Römer's determination of the speed of light involved primarily a knowledge of the diameter of Jupiter.

7. Both diffraction and refraction involve bending of light rays although for different reasons.

8. In the interference experiment pictured in Fig. 201, dark spots appear on the screen at points whose distances from S_1 and S_2 differ by a whole number of half wave lengths.

9. The polarized light on one side of a Polaroid disk would be expected to be somewhat less intense than the unpolarized light on the other side.

10. The name of Thomas Young is associated with one of the modern experiments the results of which do not fit the wave theory.

Chapter 31

LIGHT (Continued)

As we saw in the preceding chapter, at the present stage of our knowledge we have to think of light as leading a kind of double life. Under some conditions it behaves in a distinctly wave-like fashion, while under others its deportment is definitely the sort that one would expect from particles. In the topics that will be considered in this chapter we shall think of it exclusively in the first of the two roles.

The Electromagnetic Spectrum

We have already used the term "spectrum" in connection with visible light. It will be recalled that by the visible spectrum one means the range of wave lengths, or frequencies, running from red through orange, yellow, green, blue, indigo, and violet (ROY G BIV). But experiment shows that there are waves longer than red and others shorter than violet, which have exactly the same nature as light. It is the entire range, of which the visible portion is but a small fraction, that is called the *electromagnetic spectrum*.

Table 33 lists the groups into which this total spectrum may be divided. It should be emphasized that the wave-length limits are approximations. This means that at each division point there is an overlapping borderline range of wave lengths that might be put in either group. In connection with the column giving the number of octaves in each group, the reader should recall that an octave is defined as the interval from a given frequency to either twice or half that frequency, or from a given wave length to half or twice that wave length. Let us look briefly at certain details of each of the groups.

Long Electromagnetic Waves

The waves in this group are produced by particular types of electrical circuits, a fact which applies to all kinds of radio waves. For

TABLE 33. THE ELECTROMAGNETIC SPECTRUM SHOWING APPROXIMATE WAVE LENGTHS OF THE VARIOUS TYPES OF WAVES

Kind of Waves	*Approximate Wave Length Range — English Units*	*Approximate Wave Length Range — Metric Units*	*Number of Octaves*
Long Electro-magnetic Waves	Hundreds of miles to 10 miles	Hundreds of kilometers to 17 kilometers	5 or 6
Assigned Radio Range	10 miles to 29 inches	17 kilometers to 75 centimeters	14
Ultra-short Radio Waves	29 in. to .01 in.	75 cm. to .025 cm.	12
Infra-red or Heat Rays	.01 in. to .000,028 in.	.025 cm. to .000,07 cm.	9
Visible Light	.000,028 in. to .000,014 in.	.000,07 cm. to .000,035 cm.	1
Ultra-violet Rays	.000,014 in. to .000,000,5 in.	.000,035 cm. to .000,001,3 cm.	5
X-rays	.000,000,5 in. to .000,000,000,4 in.	.000,001,3 cm. to .000,000,001 cm.	10
Gamma Rays	.000,000,000,4 in. to .000,000,000,04 in.	.000,000,001 cm. to .000,000,000,1 cm.	4

the sake of the table we have rather arbitrarily placed the upper limit of wave lengths at hundreds of miles. Theoretically, there is no length which can be said to constitute an absolute maximum. The longer ones listed here have no particular practical applications at the present time. However, those ranging in wave length from something like 12 to 19 miles, which is about 20 to 30 kilometers, are used in some radio telegraph work.

Assigned Radio Range

This is the portion of the radio spectrum that is allocated for specific purposes by the Federal Communications Commission. Wave lengths from about 17 kilometers down to 545 meters are given over largely to government communication including ship, aircraft, and coastal telegraph. This is a frequency range of from some 17.6 kilocycles up to 550 kilocycles. The part of the band from 545 meters down to about 189 meters constitutes the range assigned for commercial broadcasting. This is the frequency band from 550 kilocycles to 1,590 kilocycles as shown on the standard radio dial. To illustrate the change from frequency to wave length, let us consider a station that broadcasts on a frequency of 550 kilocycles. In radio frequencies the term "cycles" may be taken as synonymous with "vibrations per second." Since the prefix "kilo" means "thousand," 550 kilocycles is equivalent to 550,000 vibrations per second. Now all electromagnet vibrations travel with the speed of light, namely, 3×10^{10} centimeters per second. Substituting the above values in the relation developed in Chapter 28, which was,

$$Speed = Wave\ length \times Frequency,$$

we get,

$$(3 \times 10^{10}\ cm./sec.) = (Wave\ length\ in\ cm.) \times (550{,}000\ vib./sec.)$$

or

$$Wave\ length\ in\ cm. = \frac{30{,}000{,}000{,}000\ cm./sec.}{550{,}000\ vib./sec.}$$

$$= 54{,}545\ centimeters = 545 + meters.$$

In order to fix this relationship in mind the reader should take the wave lengths and frequencies of several radio stations as they are given in the newspaper, change them to centimeters and vibrations per second respectively, and see whether when the two quantities are multiplied together, the product equals the speed of light.

Just below the commercial broadcast band is the range of wave lengths from 189 meters to 174 meters, including frequencies between 1,590 and 1,716 kilocycles. This band is allocated for some government work and for use by certain marine and police stations.

Below this comes an amateur band which goes down to a wave length of 150 meters, or 2,000 kilocycles. Another amateur band covers the range from 3,500 kilocycles to 4,000 kilocycles, with still another from 7,000 kilocycles to 7,300 kilocycles. These are wavelength ranges around 78 meters and 42 meters respectively. Foreign and domestic short-wave communication uses a series of bands at 49, 31, 25, 19, 16, and 13 meters. Short radio waves have an advantage in that they can be focused and confined to a narrow beam to a much greater extent than can the longer waves. Instead of broadcasting with the energy spreading out in all directions, short-wave stations can " narrowcast " in such a manner that most of the energy is directed toward the desired objective. In this way a maximum of distance is achieved with a minimum of power.

Wave lengths from 10 meters down to about 75 centimeters, that is, the frequency range 30,000 to 400,000 kilocycles, are assigned for use by experimental stations, certain government departments, television stations, and some police communication outfits.

Some of the uses, other than for radio work, that are made of waves from 75 centimeters up to several meters in length will be mentioned in the next section since the two bands are involved.

Ultra-short Radio Waves

A considerable amount of experimental work has been done within recent years on electromagnetic waves in this range. They have valuable applications in two-way radio telephone communication over relatively short distances. Such a telephone system formerly operated between France and England using 18-centimeter waves. Waves down around 5 centimeters in length have been employed to some extent in two-way police communication. As mentioned above, the high directional characteristics of these waves make it possible to use very small power output.

There are a number of applications for waves in the upper part of this band and the lower portion of the preceding one which have nothing to do with radio communication. Wave lengths of several meters are used to kill various types of organisms, such as weevils or worms, when they are in the egg or larvae stage in grain. Artificial fevers, which have proven so valuable in the treatment of certain dis-

eases, are induced by subjecting the patient to very short radio waves. So-called " diathermy " machines are really short-wave transmitters.

The lowest part of the range, as we have stated it in the table, has no practical uses at the present time.

Infra-red or Heat Rays

It is a familiar fact that when any object is at a higher temperature than its surroundings, it radiates heat. This heat can be felt, although

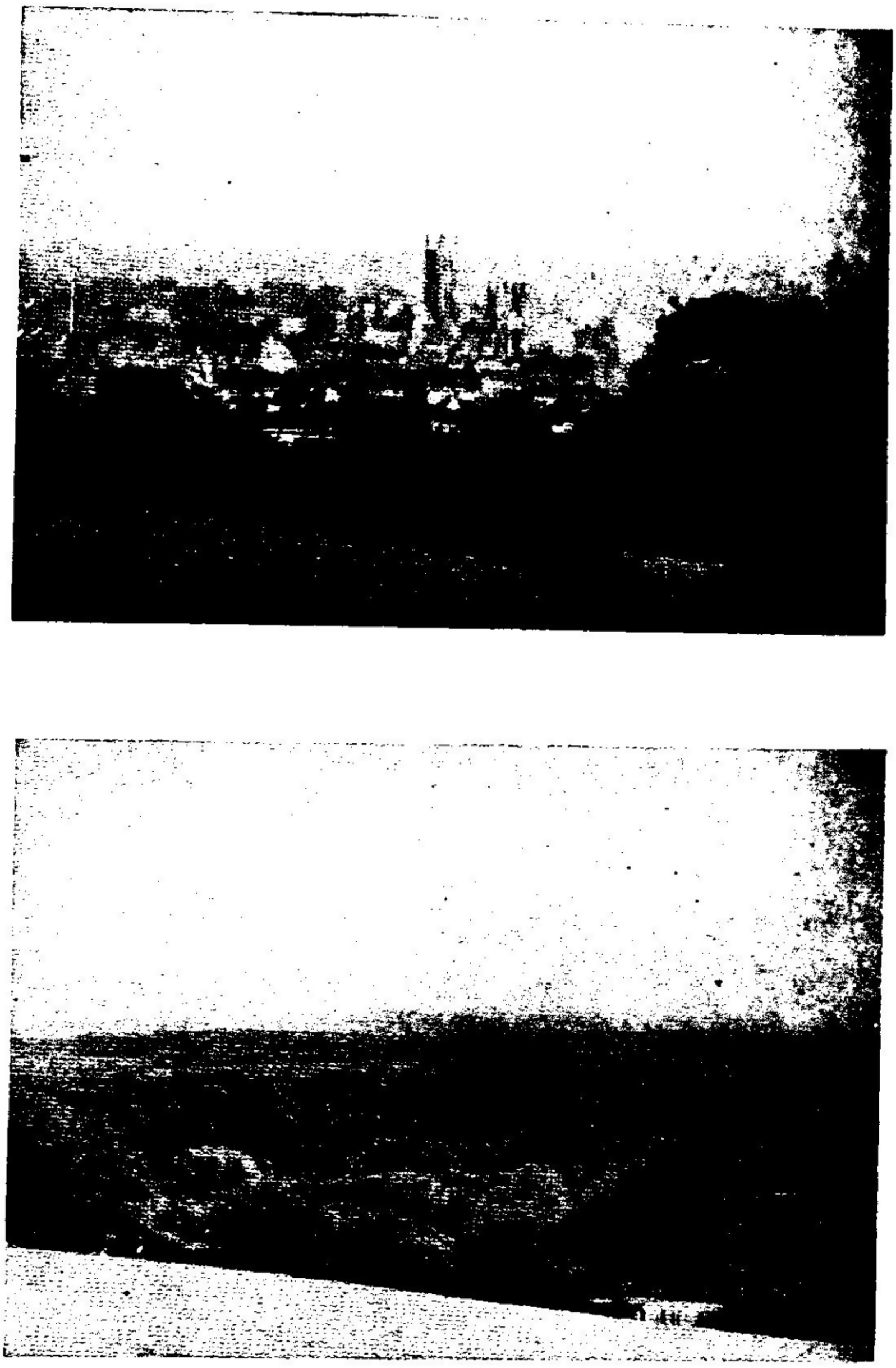


FIG. 207. Photographs of a view under clear conditions on ordinary film (above) and infra-red sensitive film (below). (Photos: Messrs. Ilford, Ltd.)

the temperature of the radiating material may be far below that at which it would become incandescent and emit light. Experiment shows that the radiation is composed of a range of wave lengths, all of which are shorter than radio waves but longer than red light. Although infra-red rays are not visible to us, photographic films and plates are made today which are sensitive to them, and it is in connection with such photography that some of the most interesting applications of these waves take place. For one thing, it is possible to photograph an object by the " light " of an electric iron or a stove in a room that appears totally dark to the eye. A somewhat more practical application of infra-red photography concerns the photographing of distant objects. It happens that as light goes through the atmosphere there is a tendency for the rays to be scattered, that is, reflected at random, by particles of dust or mist that are present. But the amount of scattering becomes less, the longer the wave length. For this reason, infra-red rays penetrate better than does visible light. The comparative photographs shown in Fig. 207 illustrate this fact. One was taken on ordinary film and the other on infra-red sensitive film. In such pictures the green of trees and grass photographs a kind of silvery color which looks very much as though the scene were bathed in moonlight.

An application of infra-red in modern warfare concerns the detection of the presence and location of enemy aircraft. Detectors sensitive to the heat from the engines of airplanes make possible the calculation of their approximate location even though they are quite invisible, in the ordinary sense of that term.

Visible Light

The characteristics of visible light were considered at some length in the preceding chapter and most of its applications are pretty well known. The number of octaves included in the visible range is worth noting. Out of the approximately 60 octaves represented in the whole electromagnetic spectrum, we can see only one. Thus, our eyes are sensitive to a much smaller range of light frequencies than are our ears to sound frequencies. The audible sound range comprises some 9 or 10 octaves.

Ultra-violet Light

The group of electromagnetic radiations having wave lengths just too short to be visible is called ultra-violet light. These rays have received a great deal of attention in recent years in connection with their beneficial physiological effects. The lack of vitamin D can be compensated for, to a large extent, by the use of ultra-violet radiation. If young rats are fed on a diet that is adequate in every respect except that it contains no vitamin D and are kept in the dark, they exhibit symptoms of rickets in a short time. On the other hand, as little as 20 minutes a day of sunshine will keep the disease from developing. That it is the ultra-violet rays that produce the results is proven by the fact that sunlight coming through glass which absorbs these particular wave lengths will not prevent rickets. Experiments have shown that the reactions of human beings are similar. Of course, the physiological effects of ultra-violet light are sometimes quite the opposite of beneficial, as overindulgence in sun bathing demonstrates. The fact that these rays are not visible makes it possible to develop a severe case of sunburn on a day when one feels quite safe because the sun appears to be hidden behind thin clouds.

An application of ultra-violet rays that has acquired considerable importance in the last few years is in connection with fluorescent lighting, a subject that is discussed more fully in Chapter 33.

X-rays

These rays are generated when electrons traveling at a high rate of speed strike a target of some heavy metal such as platinum or tungsten. An X-ray tube is a device which, first, provides electrons in the same manner as that described previously for the Crookes tube, then speeds them up, and finally, allows them to hit a piece of heavy metal. Note that the electrons are not the X-rays. Someone has suggested the analogy of machine gun bullets being fired against a metal target and setting up sound waves when they strike. The electrons in the X-ray tube are analogous to the bullets and the X-ray radiation to the sound waves. The principle applications of X-rays result from their comparatively great ability to penetrate matter. This property, in addition to the destructive effect the rays have on living tissue, gives them their value for the treatment of a disease like cancer. It is this

same penetrating power that makes X-ray photographs possible. In the radiograph of the human hand shown in Fig. 208 the X-rays penetrated the flesh much more easily than they could the bone structure, producing a silhouette of the bones on the photographic film. Note that the finger ring essentially stopped any rays at all from getting through. The use of extremely short exposure times in taking X-ray photographs represents one of the more recent developments in the field of X-ray photography. Tubes have been developed which make millionth-of-a-second exposures possible. The technique permits the X-ray study of moving parts of machinery such as rotating shafts and fly wheels. In this way flaws can be detected which are not apparent when the member is at rest.



FIG. 208. X-ray photograph of a human hand. (Photo by Dr. Whateley Davidson.)

Gamma Rays

Gamma rays, we have already seen, represent one of the three types of radioactive radiation, the other two being electrons and alpha particles. Being of shorter wave length than X-rays, they are even more penetrating. They find uses similar to those mentioned for X-rays in the treatment of diseases, being employed in cases where the trouble is too deep-seated to be reached by the longer rays.

Gamma ray photographs can be used in the study of the internal structure of large pieces of metal. In applying this technique a capsule containing about half a gram of radium sulfate — $12,000 worth — is placed on the inside of a steel casting. Gamma rays emitted by this compound will penetrate several inches of steel and register on the film which is strapped around the outside. An example of such a

gamma ray photograph of steel is shown in Fig. 209. The dark spots indicate air bubbles and impurities. From pictures of this sort metallurgical engineers are able to detect flaws and prescribe corrective treatment for the metal.

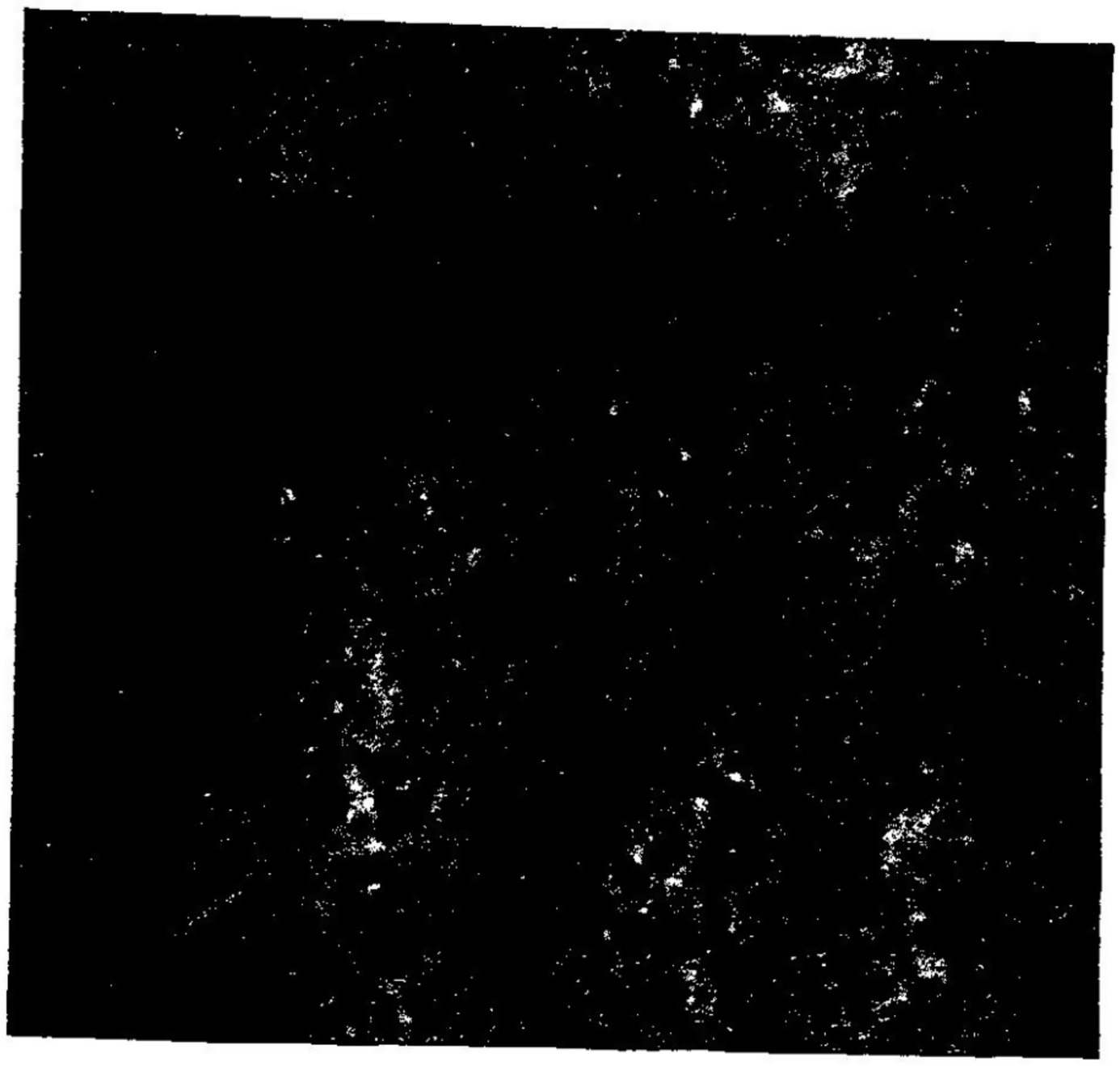


FIG. 209. Gamma ray photograph of a piece of steel. The dark spots are internal air bubbles and impurities. (Westinghouse Electric and Manufacturing Co.)

Cosmic Rays

Cosmic rays have received an enormous amount of newspaper and magazine attention in recent years. As a matter of fact, their nature has not been established as satisfactorily as has that of the other types of radiation just considered. However, cosmic rays are known to be more penetrating than gamma rays and if, as some of the evidence suggests, they should be included in the electromagnetic spectrum, then they occupy a position at the bottom of our table. This means that their wave lengths would be less than the four hundred billionths of an inch that marks the approximate lower limit of the gamma-ray range.

THE PHYSICAL BASIS OF SEEING

Since the action of the human eye resembles in many respects that of a camera, we shall begin this part of the discussion by comparing these two instruments.

Comparison of the Human Eye with a Camera

The Camera. — The diagram in Fig. 210 is a cross-section of a bellows-type camera and shows the formation of an image on the film. The point at which an image is located back of the lens varies with the distance that the object being photographed is in front of it. For example, the rays of light from an object that is very far away come

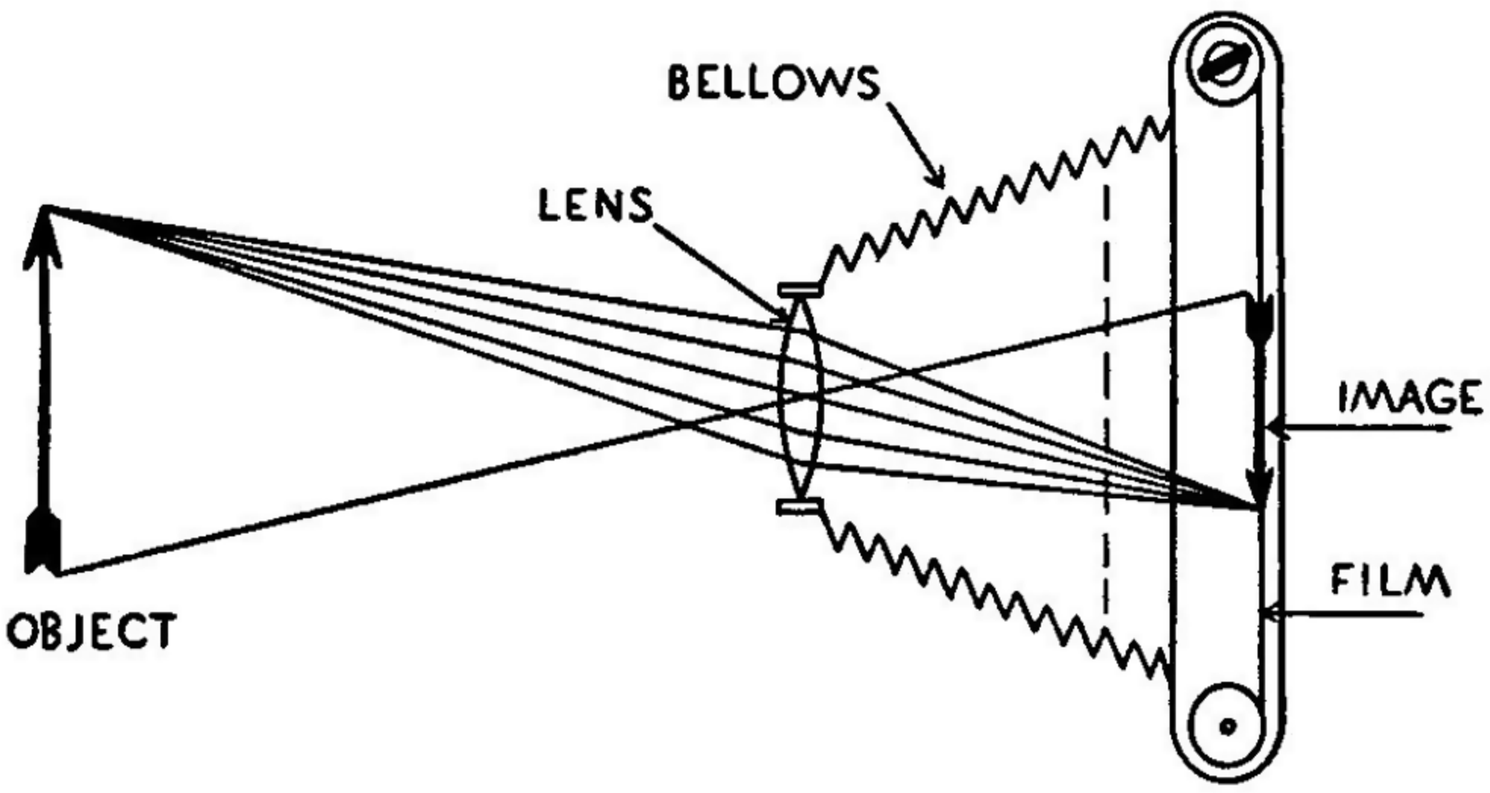


FIG. 210. Diagram illustrating the formation of an image on the film in a camera. (After White.)

to a focus at some such place as that indicated by the dotted line. To photograph a distant object it would be necessary to move the lens back until the dotted line coincided with the film. For objects nearer the camera than the arrow, the lens would have to be moved farther out. This process is called *focusing*. The amount of the variation of image location with change in position of the object is less for small lens openings than for large. As the reader may have discovered from experience the more a camera lens is "stopped down," the greater is the range of distance over which objects are in good focus. So-called fixed-focus instruments, like box cameras, are made with a small lens opening. In modern small cameras of the "candid" type, focusing is accomplished by screwing the lens in or out of the mounting, instead of by sliding a bellows back and forth. A single converging lens of the sort shown in the diagram is used only in very cheap cameras, and is subject to quite a number of lens defects. These can largely be eliminated by combining a number of lenses. Most good cameras contain from three to five separate lenses in the complete combination. Note that an inverted image is formed on the film.

The Eye. — Fig. 211 represents an arrangement for the eye analogous to that in the preceding diagram for the camera. The lens of the eye corresponds to the lens of the camera. The retina is the sensitive screen which is analogous to the film. As in the case of the camera, an inverted image of the object is formed on the retina, even though the brain interprets this image as right side up. It is possible

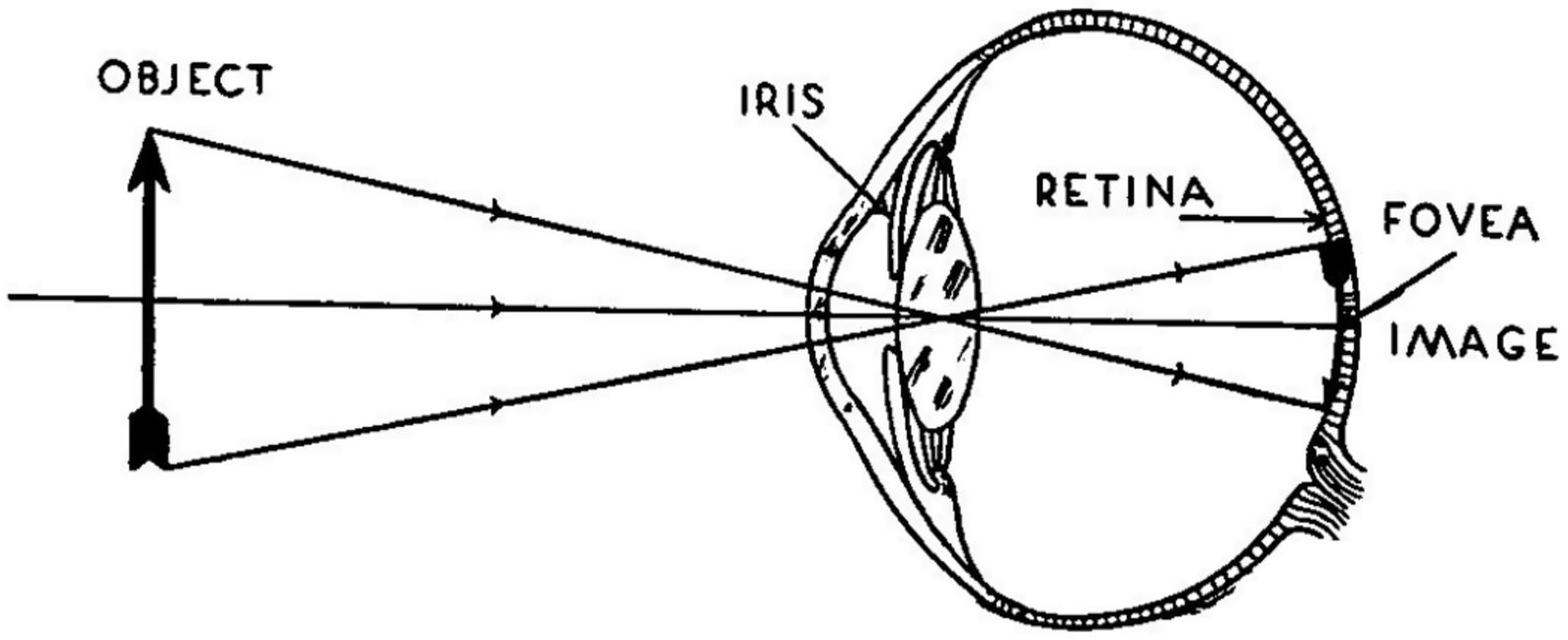


FIG. 211. Diagram illustrating the formation of an image on the retina of the eye. (After White.)

to demonstrate this fact by an arrangement which is known to throw an upright image on the retina. When this is done the subject sees the object inverted. The iris is a flexible diaphragm which varies the amount of light entering the eye just as the stop or diaphragm does in the camera.

The eye differs from the camera in the manner in which adjustment is made for objects at different distances in front of the lens. Instead of moving the lens back and forth, its curvature is varied by a system of muscles and ligaments which surround it. When one looks at nearby objects certain muscles squeeze the lens and make it bulge out, bringing the point of focus closer to it. For distant vision other muscles pull at the edges making the lens thinner and moving the focus point back. In this way the image is kept on the retina regardless of the location of the object. This process, which accomplishes the same result as focusing in the camera, is called *accommodation*. When the ability for accommodation decreases, one has to add supplementary lenses in the form of spectacles to aid in the maintenance of good vision.

It is worth noting that fish bring objects to focus on the retina of their eyes in the same manner as that described for the camera, that is, by moving the lens in and out.

Aids for Defective Eyes

If, for any reason, the lens of the eye hardens or the muscles grow weak, the accommodation process becomes difficult. When the lens happens to grow hard in a shape that is too thin, the condition known as far-sightedness, or hypermetropia, develops. In the far-sighted eye, the image forms back of the retina, as shown by the dotted line of Fig. 212(a). Since the rays must focus on the retina for clear vision, the remedy is to provide a converging spectacle lens that aids in bringing them to a focus somewhat more quickly.

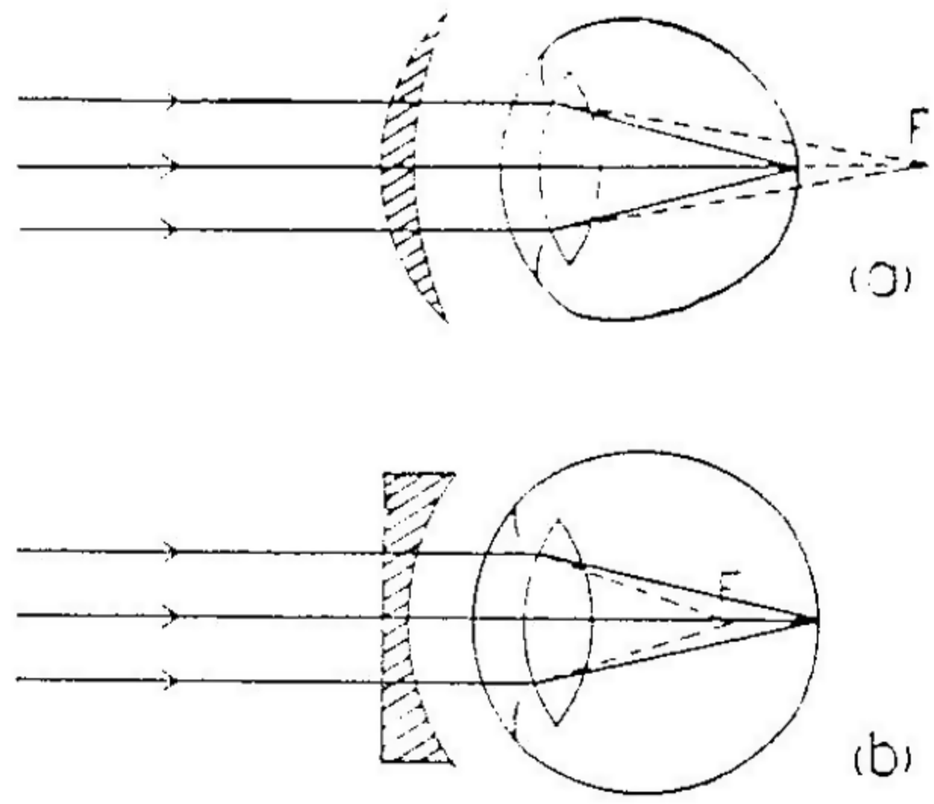


FIG. 212. Diagrams showing correction of (a) far-sightedness by the use of a converging lens and (b) near-sightedness by the use of diverging lens.

If the lens of the eye is too thick, the rays focus in front of the retina and the individual is said to be near-sighted. This condition and the remedy are shown in Fig. 212(b). For correction, a diverging lens is supplied which prevents the light from focusing quite so quickly, and moves the image back to the retina. Near-sightedness is also known as myopia.

A third common eye defect is astigmatism. This results when the lens of the eye is not of uniform curvature in all directions. Fig. 213 is a test chart for this defect. The remedy is to provide spectacle lenses which also are unequally curved but in a manner that will just counteract the defects of the eye lens.

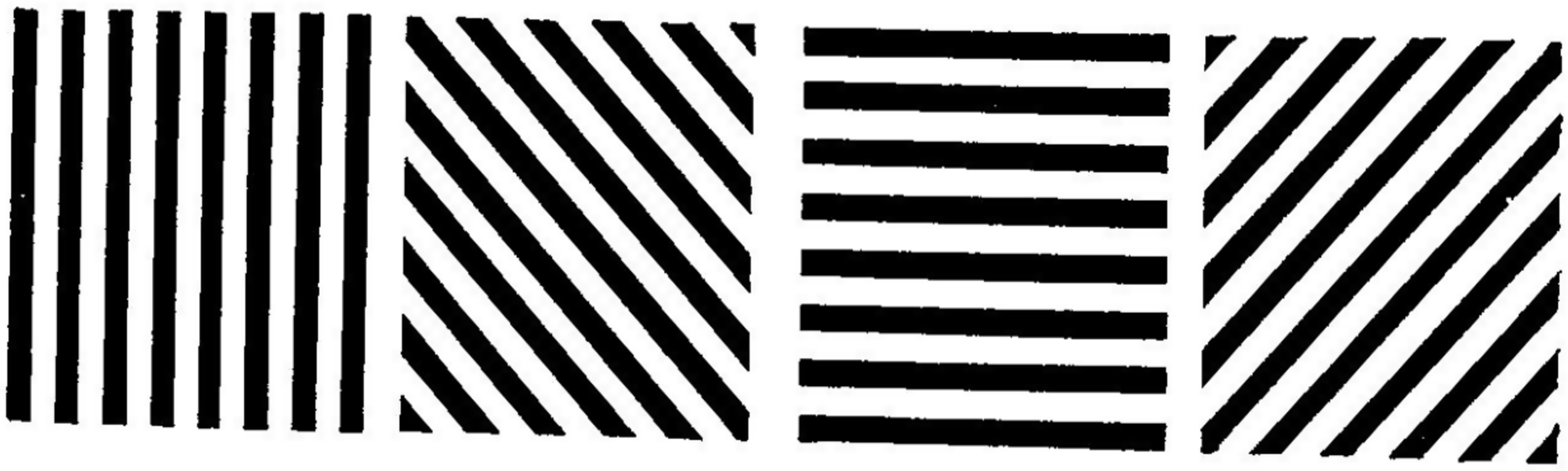


FIG. 213. Test Chart for astigmatism. The lines should all appear of equal blackness to the normal eye.

Some Aids for the Normal Eye

There are a number of special cases in which man finds it desirable to supplement the normal eye with optical instruments. The astronomical telescope, two types of which were described back in Chapter 5 is one such device. When the refracting type of this instrument is

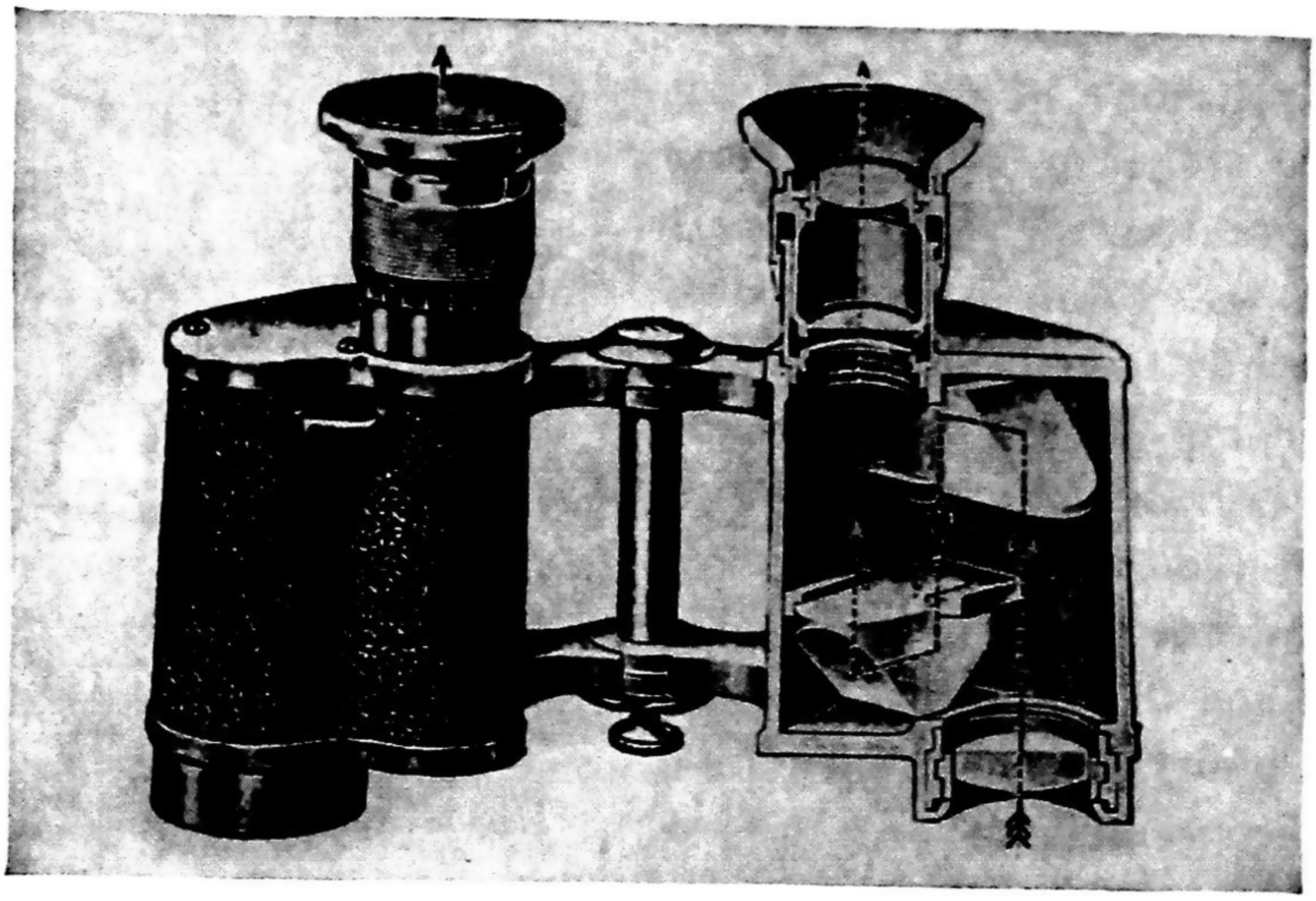


FIG. 214. Diagram of prism binoculars showing lenses and total reflecting prisms. (Courtesy of the Central Scientific Co.)

modified for use on the earth we call it a spy glass or terrestrial telescope. The astronomical telescope, it will be recalled, contained two converging lenses and produced an inverted image, a feature that would be objectionable in a device to be used for viewing earthly objects. Consequently, a third converging lens is added which has the function of reinverting the image, so that the final one as seen by the observer is erect, or right side up. In the opera glass, the lens system consists of a converging objective lens with a diverging lens being used for the eyepiece. When properly located with respect to each other, such a combination produces an erect image. Galileo's original telescope had this sort of lens arrangement. In prism binoculars the production of an image that is right side up is accomplished in still another way. The prisms are so arranged as to " totally reflect " the light, its path being shown in Fig. 214. One result of the reflection

by the prisms is to produce an erect image, although both the objective and eyepiece lenses are of the converging type. A second function of the prisms is that of increasing the light path. For example, an instrument of the same magnifying power as the one shown, but having single tubes like field glasses, would have to be approximately three times as long as the prism type instrument.

The microscope in its several forms constitutes another valuable optical instrument. It was invented about the same time as the telescope, both of them being outgrowths of the spectacle industry.

A simple microscope, such as a reading glass or pocket magnifier, consists of a single converging lens. By placing the object to be viewed closer to the lens than the point at which parallel rays of light focus, the observer sees a magnified image. The highest practical magnification that can be obtained in air with a simple microscope is about 400. The compound microscope, invented by Galileo, resembles the astronomical telescope in that it has an eyepiece and an objective lens, both of which are of the converging type. The principle difference in the optical features of the two lies in the fact that lenses of comparatively short focal length are used in the microscope. In modern high-grade instruments the objective itself may consist of a train of eight or ten lenses while two or three are combined in the eyepiece, thus eliminating certain optical defects that are inherent in single lenses. With proper illumination, magnifications of 1,000 diameters are easily obtained. For magnifications greater than 1,000 a limitation due to the wave length of the light with which the object is illuminated makes itself felt. Because of diffraction effects, objects or details of objects smaller than about half the wave length of the light being used cannot be seen clearly. So the shorter the wave length of the light, the smaller the object that can be observed. Using ultra-violet light and photographing the image, the upper magnification limit is approximately 2,500 diameters. It should be noted that this is not a limitation caused by lack of refinement in apparatus. It is a fundamental one based on properties of matter. Rays shorter than ultra-violet light, that is, X-rays and gamma rays, cannot be used because they cannot be sufficiently refracted by lenses to be brought to a focus.

It will be recalled that a so-called ultra-microscope was described in Chapter 18 in connection with the discussion of the properties of

colloids. In it, however, no details of the tiny particles can be seen but only the fact that particles are present.

A relatively new device called the " electron microscope " permits an entirely different approach to the problem of high magnification. The ordinary optical instruments described above function as they do because of the refraction or bending of light rays when they pass through glass lenses. Now electrons cannot be refracted by glass but their direction of travel can be changed, as we have seen, by passing them through electric or magnetic fields. Also, they register on photo-

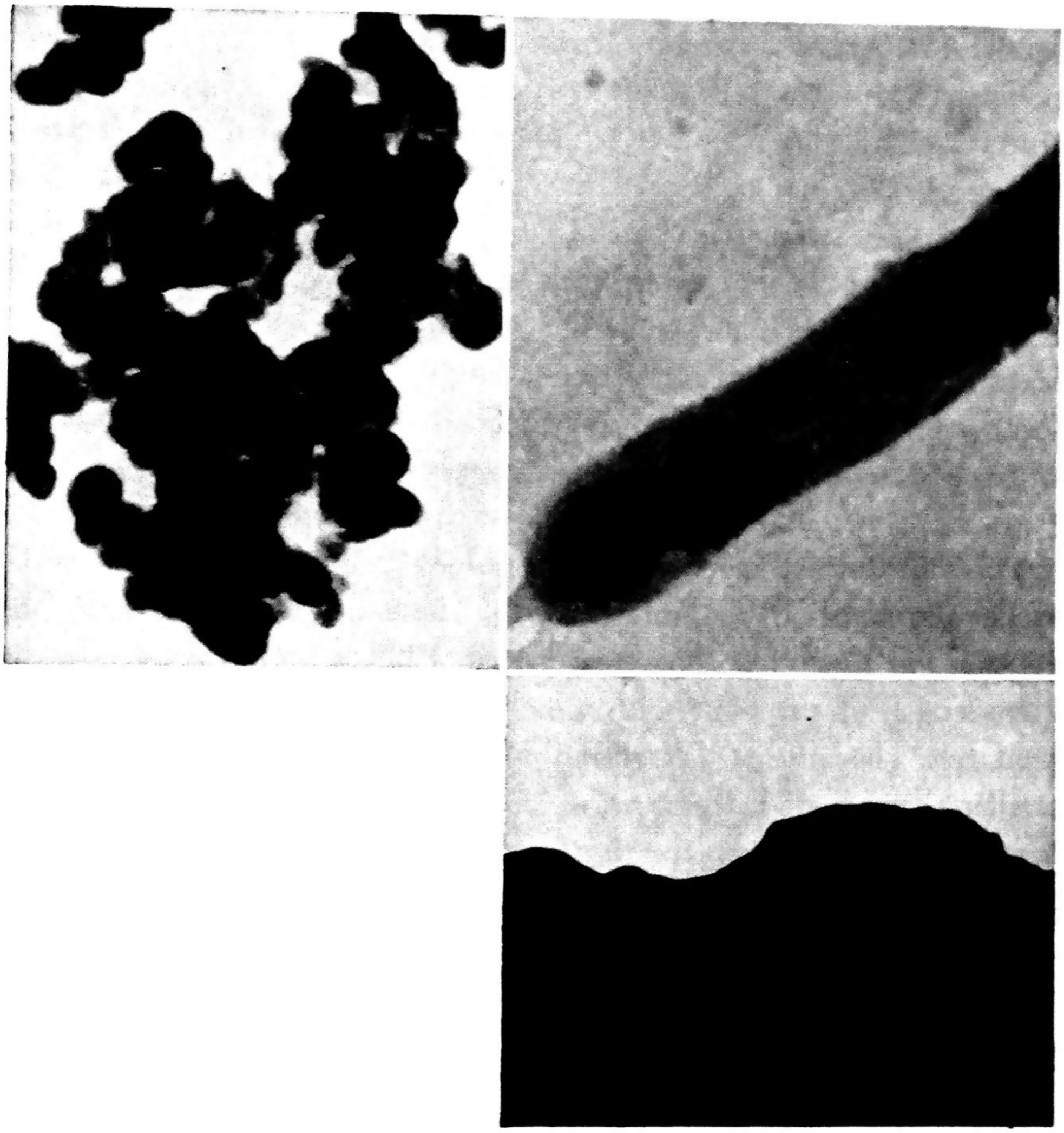


FIG. 215. Electron microscope photographs taken at University of Toronto. (a) Grains of carbon black magnified 100,000 diameters. (b) Diphtheria bacillus magnified 35,000 diameters. (c) Edge of a new safety razor blade magnified 24,000 diameters. (Courtesy Scientific American Magazine.)

graphic films when they strike them. Consequently, in the electron microscope, instead of light going through a series of lenses, electrons move through a series of three magnetic coils. Their motions are such that ones leaving the same point on the object will come together again, that is, focus, at a corresponding point on the film to produce that portion of the image. With these devices, magnifications many times that possible in the compound microscope are obtained. The pictures shown in Fig. 215 are three examples of electron microscope photographs. The one of carbon grains in lampblack is particularly noteworthy. The individual grains, some of which are about one eight-millionth of an inch in diameter, are completely invisible in even an ultra-microscope.

Color

A number of references have already been made to color both in this and in preceding chapters. Perhaps the basic point to be recognized about color is that it represents the interpretation that man's eyes and brain make of different wave lengths, or frequencies, of light. Color is always associated with the way we see things. It is our purpose in this section to bring together the facts already mentioned and a number of additional ones into a unit which covers the important features of the nature and manifestations of color.

The Nature of Color

Newton's Experiments. — The series of experiments by which Newton demonstrated that white light is composed of all the colors of the visible spectrum are among the most important ever performed in science. The first of the group, which was mentioned briefly back in Chapter 5, is pictured again in (a) of Fig. 216. As we saw in the earlier reference, when white light enters a prism, there emerges from the other side of it the band of colors which are called variously the seven colors of the rainbow, the continuous spectrum, or the visible spectrum. This particular effect had been observed before Newton's performance of the experiment, but had in general been explained by assuming that the prism effected some fundamental change in the nature of the white light, thus actually creating the colors. Two experiments performed by Newton to test this theory are illustrated in

(b) and (c) of the same figure. In (b) a second prism was placed behind the first. Between the two there appeared the continuous spectrum which, after going through the other prism, changed back to white light. The demonstration shown in (c) consisted in picking out one color from the seven and allowing it only to pass through a second prism. Newton reasoned that if the prism had the property of changing light, then some different color should emerge from the other side. Actually, no change at all was made in the single color by prism number two. These facts proved to Newton's satisfaction that white light itself is a combination of the various colors and that the prism simply served to separate them. The rest of the scientific world, however, did not readily accept this conclusion, the announcement of which aroused such bitter criticism of Newton's work that for a time he considered giving up scientific work entirely.

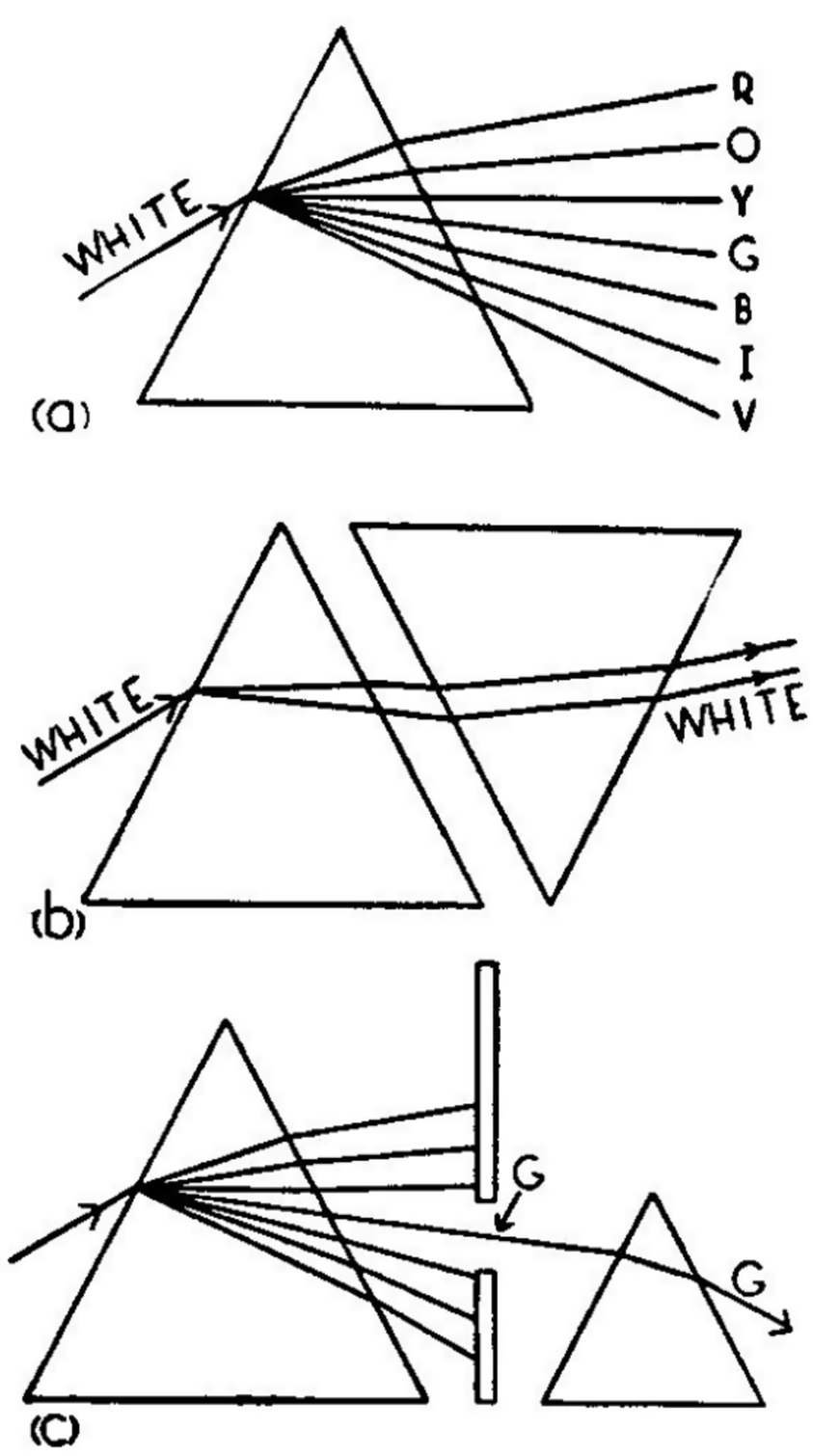


FIG. 216. Newton's experiments which proved that white light consists of all colors. (a) Effect of passing white light through prism. (b) Effect of passing light through two prisms. (c) Effect of passing light of a single color through prism.

The most common natural manifestation of the splitting up of white light in this fashion is the rainbow. Note that the effect as described for the prism is a refraction phenomenon with the separation of the colors occurring because different ones are bent different amounts in going from one medium to the other. In the case of the rainbow, drops of water take the place of the glass prism. As a ray of sunlight enters the drop, refraction and splitting up of the light occurs just as at the entering edge of the prism. The effect is increased when the ray emerges again after having been partially reflected once within the drop. The so-called secondary rainbow, with the colors arranged in reverse order, results from refracted rays that emerge after two reflections within the raindrop. Halos around the sun and moon,

commonly called rings, have the same sort of cause. In this case, the refracting objects are tiny crystals of ice in the upper atmosphere.

Color of Materials. — Now we may ask just what one means by saying that a material is a certain color. Just how would one define a red object, or a blue one? Definitions of this sort may be established by the following experiments:

1. The basic definition must be one of a white object. To get this we permit the various colors of light to shine, one at a time, on a screen of the sort which we call white. Doing this we find every one of them to be faithfully reflected. So we may define a white material as one which reflects all colors or wave lengths.
2. Now suppose we permit white light to fall on a material of the sort that we call red. The results admit of two possible explanations. Since we know from Newton's experiments that all colors are present in the white light, it might be that the red material has the ability to change all these colors to red. The other possibility is that the material picks out the red portion of the white light and reflects it while absorbing all the other colors.
3. As an experiment to test these alternatives we may permit blue or green light to fall on the red material. If the first of the above possibilities is correct, then the object should still appear red. If the second is the true explanation, then the material should appear black, since everything but red is absorbed. Performing the experiment, we find the substance looks black under blue or green light. So we may define the color of any object as being that of the light which it reflects.

Mixing of Colors. — Here we are concerned with the effects of mixing together lights of different colors and pigments of different colors. In the second of Newton's experiments illustrated in Fig. 216, the right-hand prism mixed the seven colors together and produced white light again. Similarly, by rapidly rotating a so-called color disk which contains segments of the seven colors of the proper relative sizes, the sensation of almost pure white can be produced. It is not necessary to mix all seven colors of the visible spectrum in order to get white light. It can be produced by mixing blue, green, and red light. These three are called the primary colors because it is possible to pro-

duce any hue at all by mixing them in the proper proportions. This fact is used extensively in stage lighting where by varying the relative intensities of groups of red, green, and blue bulbs, resultant light of the desired color can be obtained. White light can also be obtained from various combinations of two colors. Such pairs are said to be complementary. For light of any color a complementary color can be found which, when properly added to the first, will produce white light. Examples of complementary color pairs include:

Yellow and Blue	Red and Bluish-green
Orange and Greenish-blue	Violet and Greenish-yellow

It should be noted that what we have said in this paragraph applies to the mixing of *light*.

When one mixes paints or pigments, quite different results are obtained. Let us consider yellow and blue pigment and ask what we might expect the result of mixing them to be, in the light of the previous discussion of the color of materials. Our conclusions there would lead us to argue that if yellow pigment absorbs all colors but yellow, and blue pigment all colors but blue, a mixture of the two would absorb all colors and so appear black. But the only way to be certain exactly what colors a given pigment reflects is to split up the light reflected by it. As we have seen, this can be done by passing the light through a prism. When this is done with light reflected by the yellow pigment, we find not only yellow but mixed with it some red, orange, and green. Light reflected by the blue pigment is found to contain some violet and green in addition to the blue. Thus, neither one completely absorbs green so that a mixture of the two would be expected to appear green, as it does. The same sort of thing occurs with other pigments.

The three primary pigments are red, yellow, and blue. Mixing them always tends to produce black rather than white. When they are mixed in pairs:

A red pigment + a yellow pigment produces an orange pigment;
A red pigment + a blue pigment produces a violet pigment;
A blue pigment + a yellow pigment produces a green pigment.

Color Vision

Many attempts have been made to explain color vision but so far none has been completely successful. The one that most nearly accounts for all known experimental facts is called the Young-Helmholtz theory. It was first advanced by Thomas Young, of wave theory fame, and later amplified by the German scientist, Hermann von Helmholtz. This theory postulates the existence of three kinds of minute cones in the retina of the eye, one of which produces the visual sensation of red, that is, makes us see red, another green, and the third, blue-violet. The color of an object at which we look is thought to excite one or more of these sets to varying degrees of activity. The brain then is pictured as taking the three responses and mixing them in such a way as to produce a kind of net sensation which we call the color of the object. As a result of this mixing we are able to distinguish thousands of different hues and shades although there are only the three kinds of cones.

Color-blindness, according to the Young-Helmholtz theory, results from the absence or deficiency of one or more sets of cones. The person who cannot distinguish red from green is lacking in either the red or green sensitive cones. An individual lacking all three types would see everything in shades of black and white. This latter type of color-blindness is rare but occurs occasionally.

The principal weak point of the theory is that so far no one has been able to produce any anatomical evidence of the actual existence in the eye of three different kinds of receivers.

The Production of Color

There are several ways in which colors can be produced from white light. Two of these which have already been considered are, by passing the white light through a prism, and by allowing it to be reflected from or transmitted through colored materials. In the first case, the white light separates into its constituents because of the fact that different colors are refracted differently. In the second method, all the colors except the one that is either reflected or transmitted, are subtracted from the white light by absorption.

A third source of color is *interference*. It will be recalled that in

our previous discussion of interference we found that light waves reinforced each other and produced a bright band, if either crests or troughs came together, but interfered and produced a dark band if a trough from one joined a crest from another. Such an effect results when light shines upon two pieces of glass arranged so that there is a thin wedge of air between them. Part of the light that falls on the outfit is reflected from the rear surface of the first plate and part from the front surface of the second plate. The second of these two light-wave trains combines with the first. Now the two reflection points are separated by the thickness of the air which increases gradually from the top to the bottom of the wedge. So at some points a crest reflected from the back plate will return just in time to join a crest from the front one and produce reinforcement. At points between these reinforcement bands crests from the second plate will join troughs from the first and cause interference. Thus, on the front plate there will be a series of alternating bright and dark bands, as shown. Since the location of the bands depends on wave length, the distance between them will be different for different colors of light, with red bands being separated the most and violet ones the least. Consequently, if white light is used, there will be a separation into colors. The colors seen in a soap bubble are produced in this manner. Because of gravitational attraction the bubble is thinnest at the top and thickest at the bottom, thus forming a very thin wedge. Light reflected from the two surfaces reinforces and interferes in the manner described above. Other examples include the colors in oil films on water and those seen in mother-of-pearl. It should be pointed out that the effect occurs only for very thin wedges. When the wedge is thick the bright bands of the various colors overlap to such an extent that they combine and one still sees white light.

A phenomenon known as *scattering* is also productive of color. It is simply the haphazard irregular reflection of light in all sorts of directions. For instance, on a foggy night, the tiny drops of water in the atmosphere scatter the light from automobile headlights. The part that is scattered backwards tends to blind the driver and make it impossible for him to see the road.

The bearing of the phenomenon on color results from a fact already mentioned in the discussion of infra-red radiation, namely, that

the longer the wave length of the light the less is the amount of scattering. Consider the sunlight that reaches us early in the morning or late in the afternoon. When it enters our atmosphere it is white light which means that it contains all wave lengths. But the shorter wave lengths, that is, those in the blue-violet end of the spectrum, are scattered more than the longer ones by air molecules and dust particles. This means that the light that reaches us contains a higher percentage than normal of the longer waves. As a result it appears unusually red. In the middle of the day when the sun is overhead, the light has a much shorter atmospheric path to traverse and so is subjected to less scattering. The blue appearance of the sky results from the scattered short wave lengths in the sun's rays. At high altitudes such as those attained in the various, stratosphere balloon flights, there is comparatively little atmosphere above the observer, and the sky appears not blue but a purplish black. The blue color of wood smoke is another example of a scattering effect. The carbon particles are very small and the color caused by scattering over-balances that of the pigment itself. In coal smoke, on the other hand, the carbon particles are larger and the pigment color is dominant, so that the smoke appears black.

Questions

Discussion

1. In Table 33 verify the number of octaves listed, for several of the types of electromagnetic waves.

2. Make two or three calculations of the speed of radio waves using data for different radio stations as given in the newspaper.

3. Find some practical applications, other than those mentioned in this chapter, for several of the types of electromagnetic waves.

4. Express in kilocycles the frequency of a radio station that broadcasts on a wave length of 375 meters.

5. Take the standard broadcast band as given on the dial of some particular radio set and calculate the number of octaves represented.

6. If orange or red lamps on an automobile are better than white for use in fog, why wouldn't infra-red lamps be still better?

7. Inquire what range of frequencies or wave length, beyond those of visible light, are emitted by one or two commercial sun-lamps.

8. Mention or list the features of the eye and a camera that are similar and those that are different.

9. Is making the lens of the eye thicker, analogous to moving a camera lens toward or away from the film, as far as focusing is concerned?

10. Discuss the principal eye defects with respect to cause and remedy.

11. If you wear glasses test yourself with the chart in Fig. 213 both with your glasses and without them.

12. What changes occur in the eye as one views a rapidly approaching automobile?

13. Why does a spyglass or terrestrial telescope contain three converging lenses whereas an astronomical refracting telescope contains only two?

14. Explain the total reflection of light by the prisms in prism binoculars.

15. Discuss the experiments and trace the reasoning by which Newton proved that white light is composed of the seven colors of the visible spectrum.

16. State a definition of a black material.

17. An object that appears black in white light should appear what color in yellow light? in green light? Explain.

18. See if you can find in this chapter any basis for an explanation of the fact that if one looks at a blue object for a few seconds and then at a white screen, one sees the object on the screen in yellow.

19. Explain why the results obtained from mixing complementary colors of light are different from those obtained when the same colors of pigments are mixed.

20. In the light of the discussion in this chapter explain why two pieces of cloth that appear the same shade under artificial light may be quite different shades in sunlight.

21. See if you can find in this chapter an explanation of the fact that a lake which appears blue on a clear day looks gray on a cloudy one.

22. Find out all you can about the Kodachrome process of color photography.

23. Find out something about the first studies made of color-blindness.

24. Find examples other than those given in this chapter of color effects due to interference and due to scattering.

Multiple Choice

1. The wave length of yellow light is greater than that of (one or more) (a) a WJZ broadcast, (b) the sun's rays that cause sunburn, (c) heat waves, (d) blue light, (e) X-rays.

2. The image on the film in a camera is in general — (a) right side up and larger, (b) right side up and smaller, (c) inverted and larger, (d) inverted and smaller — than the object being photographed.

3. In the human eye the focusing on the retina of light from an object is accomplished by (a) opening or closing the iris, (b) moving the lens back and forth, (c) varying the density of the material between the lens and the retina, (d) varying the color of the light, (e) varying the curvature of the lens.

4. A pure blue object illuminated with pure red light would appear (a) black, (b) blue, (c) red, (d) yellow, (e) green.

5. The number of " sound octaves " between tones with frequencies of 1,260 and 630 vibrations per second respectively is approximately the same as the number of " light octaves " between (a) long and short radio waves, (b) yellow light and gamma rays, (c) visible violet and visible red light, (d) the longest and shortest X-rays, (e) medium infra-red rays and visible red light.

True–False

1. Variations in frequency of light appear to us as variations in brightness.

2. The iris of the eye serves essentially the same function as do the stops in a camera.

3. Black and white may be said to be complementary colors.

4. One difficulty with the Young-Helmholtz theory of color vision is that no function can be found for three types of nerve endings that anatomists have found present in the retina.

5. A lens used to correct near-sightedness is of the same type (converging or diverging) as the kind used for the objective lens in a refracting telescope.

6. All visible light waves are shorter than all audible sound waves.

7. An object that appears green in green light, red in red light, and violet in violet light is white.

8. Ultra-violet light has proven the most valuable of all electromagnetic radiations for use in the photographing of distant objects.

9. If the lens in one of B's eyes should happen to be exactly like the spectacle lens that corrects A's astigmatism, then B also is astigmatic in that eye.

10. The image of the moon, as seen through a refracting astronomical telescope, is inverted.

Chapter 32

COMMUNICATION

Important Features of a Communication Method

Any mechanism by means of which ideas are transmitted between individuals is a method of communication. Outstanding among the questions which one might ask in judging the value of any specific method would be the following:

1. Is it convenient to use?
2. Is it accurate?
3. Does it provide a permanent record?
4. Does it provide high-speed communication over considerable distances?

The relative importance of these factors varies greatly with particular communication needs. In common conversation convenience is present; accuracy can be, but frequently isn't; and the other two ordinarily are not important. But if one is communicating with the fire department to inform them that the house is burning, both speed and accuracy are extremely important. On the other hand, if plans for a new house are being communicated from the architect to the builder, accuracy and permanence of record are of more consequence than speed, and so forth. It will be noted, as we go along, that the general trend in the development of modern methods of communication has been to combine all four of these factors to an increasing degree. Let us first look briefly at the earlier procedures that man perfected for communicating with his fellow man, and then at the more modern techniques.

Early Development of Communication Methods

Speech and Writing

It is quite impossible to say just when man first began to talk. It is generally agreed that for thousands of years before he developed a language, he was able to take care of his communication needs with gestures and grunts. In those days it was not only polite to point, it was necessary. Later on, perhaps when his thinking had developed to the place where he had ideas that were too complicated to be expressed by the methods of communication he had used in the good old days, he gave particular and permanent meanings to specific sounds, and so formed words. Since what people on one side of a lake or mountain called a tree would be quite different from the sound applied to it on the other side, different languages developed. There is some evidence to indicate that sounds for the names of things, or nouns, entered primitive languages some centuries before words of action or verbs.

Writing did not develop primarily as a method of communicating ideas from one person to another. Man's initial efforts in the direction of writing grew out of his need for keeping records of numbers of things, such as animals in his flocks, or periods of time. Some of his primitive methods for doing this were described in Chapter 11. Later came the desirability of keeping historical records, and the counting devices which man had developed were replaced by picture writing. This form evolved from a stage in which the picture represented a concrete object to one in which it became a symbol. The earliest records we have of this form of writing are those left by the Egyptians. The final step, that of using the symbol to stand for a sound, which essentially made it a letter, is believed to have been taken first by the Phoenicians. The fact that writing offered the possibility of accurately communicating ideas between individuals within neither sight nor sound of each other seems not to have been made use of to any important extent during its early history.

Other Methods

Man's first method of long distance communication was the runner or messenger. Among the disadvantages of this technique were its slowness and the danger of the messenger being captured or killed

by enemies. The use of domesticated animals increased the speed but did not eliminate the risk of capture. Various other methods for communicating over distances greater than the voice would carry have been employed at different times by many peoples, and include smoke and fire signals, drum beats, primitive equivalents of wig-wagging, flaming arrows, and reflection of sunlight from polished surfaces.

Modern Communication Methods

In the growth of modern methods of communication, the principal emphasis has been on speed in one form or another. Developments in the field of printing have constantly tended toward the production of more finished copies more quickly. At the same time, cost has decreased and quality of the work has increased. An appraisal of the quality of all the material printed, fortunately, is not part of the business of this book. In the telegraph, telephone, and radio, almost instantaneous communication over long distance has been achieved, first for signals, then for voice, and finally for voice and music. The development of wirephoto and radiophoto have added photographs to the material that can thus be transmitted, while television enlarges the scope still farther to include the transmission of pictures of events while they are occurring.

Printing

Before 1450. — The art of printing, that is, of impressing figures, pictures, letters, and words, was practised in some forms hundreds of years before the fifteenth century date that is often given as the time of its invention. Block printing of initials and the pre-Christian-era equivalent of trade-marks was practised in Babylonia and Rome. In the Orient, printing from wooden blocks can be traced back to the sixth century in China and the eighth century in Japan. In fact, it is said that the Chinese printed with movable type made out of clay as early as the eleventh century. Block printing on cloth and parchment was practised in Europe as early as the twelfth century, although in the middle of the fourteenth century all English books, proclamations, and documents were still written by hand. From about 1350 to 1450 block printing on paper came into European use in the

printing of books. In this process an entire page was carved on a single wooden block.

Improvement in Type since 1450. — The logical step after block printing was the development of movable type. It would seem that no one could work with wooden blocks representing entire pages very long, before thinking either of slicing the block into individual letters or of carving out separate letters to begin with, so they could be used more than once. The first actual use of such type occurred around 1450, and almost ever since that time a controversy has been in progress over exactly whose invention it was. The two principal contenders for the honor have been John Gutenberg of Mainz, Germany, and Lourens Coster of Haarlem, Holland. Although the former has received the most publicity and is often called the inventor of printing, the Encyclopædia Britannica [1] concludes that Coster's claim is best supported by the evidence. It has been rather commonly assumed that the first movable types were carved out of wood. If so, the period of their use apparently was rather short because all the earliest works of typography that have been preserved are made from cast metal type. Type metal that was an alloy of lead, tin, and antimony, as is that used today, began to be employed rather early in printing history.

The next improvements were automatic type-casting and type-setting. The former has been in general use since the middle of the nineteenth century, the first successful machine for accomplishing it having been developed around 1836. Automatic type-setting proved much more difficult to achieve. The principal difficulties were in devising a machine that would " justify " the lines, that is, make them even along the right-hand edge of the column, and that would distribute the type after printing. As long as these operations had to be performed by hand, little time was saved by a machine that merely set the type. The problem was solved in 1886 with the invention of the linotype machine by Otto Mergenthaler. This device combines casting and setting. An operator, working at a keyboard somewhat like that of a typewriter, releases individual letter molds, called matrices, which arrange themselves to form the mold for the " line of type " for which the machine is named. Molten type metal then flows into this mold and casts the entire line at once. The linotype machine has

[1] *Encyclopædia Britannica*, 9th Ed., Vol. 23, p. 692.

speeded up the production of printed matter enormously, because an operator can set up more type in an hour than a hand compositor could in a day. All newspapers and almost all magazines, in the United States, are set by linotype today. This is also true of all but the most expensive books.

Stereotyping and Electrotyping. — It is interesting to note that in the processes of stereotyping and electrotyping we find a kind of return to block printing, in that an entire page may be printed from a single block. In the early days of movable type, one of the difficulties encountered by the printer was that unless he had a large number of fonts of every kind of type, he had to complete the entire printing on each job before he could start setting up the next one. This was one of several factors which led to the practice of making a plaster of Paris mold of each page as it was set up. From this a metal cast of that page could be made. In the process called stereotyping, essentially the same procedure is followed except that the mold is made from papier-mâché, instead of plaster of Paris. As used today the process is particularly valuable in cases where a single advertisement or other unit of printed matter is to be run by a large number of publications. Each is sent a papier-mâché mold or mat from which that particular paper makes its own cast.

In electrotyping also a plate is made which may be an entire page of the material to be printed. This process, which is an electrolytic one, was described in some detail in Chapter 25, page 478.

The Telegraph

The telegraph, with its ability to transmit signals almost instantaneously over great distances, was the first important high-speed method of communication. As has been the case with several valuable scientific developments, people in widely separated parts of the world began to work on it at about the same time. It was around 1832 that the successful young American portrait painter, Samuel Morse, became interested in the possibility of using electricity to transmit some kind of recognizable signal from one place to another. At almost exactly the same time, two Englishmen, a scientist named Wheatstone and an army officer named Cooke, began joint work on the same problem. The results of both sets of experiments were announced to the world in 1837. The mechanisms used, however, differed considerably.

The English investigators, making use of the fact that a current-carrying wire has a magnetic field about it, employed the force of that field to move a magnetic needle. By varying the current, and therefore the field, the needle was made to point at letters on a dial and so spell out the message. Since this mechanism proved to be very inferior to the one developed by Morse, it was never widely used.

Samuel Morse, in developing his telegraph, made use of the electromagnet which had been invented about 1824 by Joseph Henry. A simple electromagnet, as described previously in Chapter 25, can be made by winding wire around a bar of some magnetic material such as soft iron. The important point about it is that when current flows in the wire the bar becomes magnetized, and when the current stops the bar loses its magnetic properties.

The principle of the telegraph, as far as its essential features are concerned, is quite simple. The transmitting device, or key, is simply a switch. The receiving apparatus, called a sounder, consists of an electromagnet above which a metal bar is held by a spring. When the transmitter key is closed, an electric current flows through the circuit and the receiver attracts the metal bar to itself. When the bar strikes the magnet it makes a sharp click. Lifting the key opens the circuit, and at the receiver the spring pulls the bar back up again. In early Morse receivers a magnetically controlled pencil and paper device wrote down the incoming dots and dashes. This was replaced by the modern sounder when it was found that operators could catch the signals from the sound of the machine just as accurately as from the visual record and with less trouble. The simple telegraph circuit as just described would carry only one message at a time. Among the important improvements that have been made in the telegraph since its invention are the developments that permit what is known as multiplex telegraphy. As a result of the work of Thomas Edison and others it is possible now to send as many as 12 messages over a single line at the same time.

The Telephone

Within a very few years after Morse and others had demonstrated the possibility of transmitting signals electrically, a number of inventive minds turned to the problem of doing the same thing with actual speech. Several experimenters achieved a little success in the decade

from 1850 to 1860, but it remained for Alexander Graham Bell to invent the device that was the forerunner of the modern telephone. Bell had a considerable background of knowledge in speech and sound, acquired while he was teaching in schools for deaf-mutes. Among the experiments which he carried out while he was engaged in this profession was one which demonstrated that a person's eardrum vibrates when sound waves strike it. This led him to consider the possibility

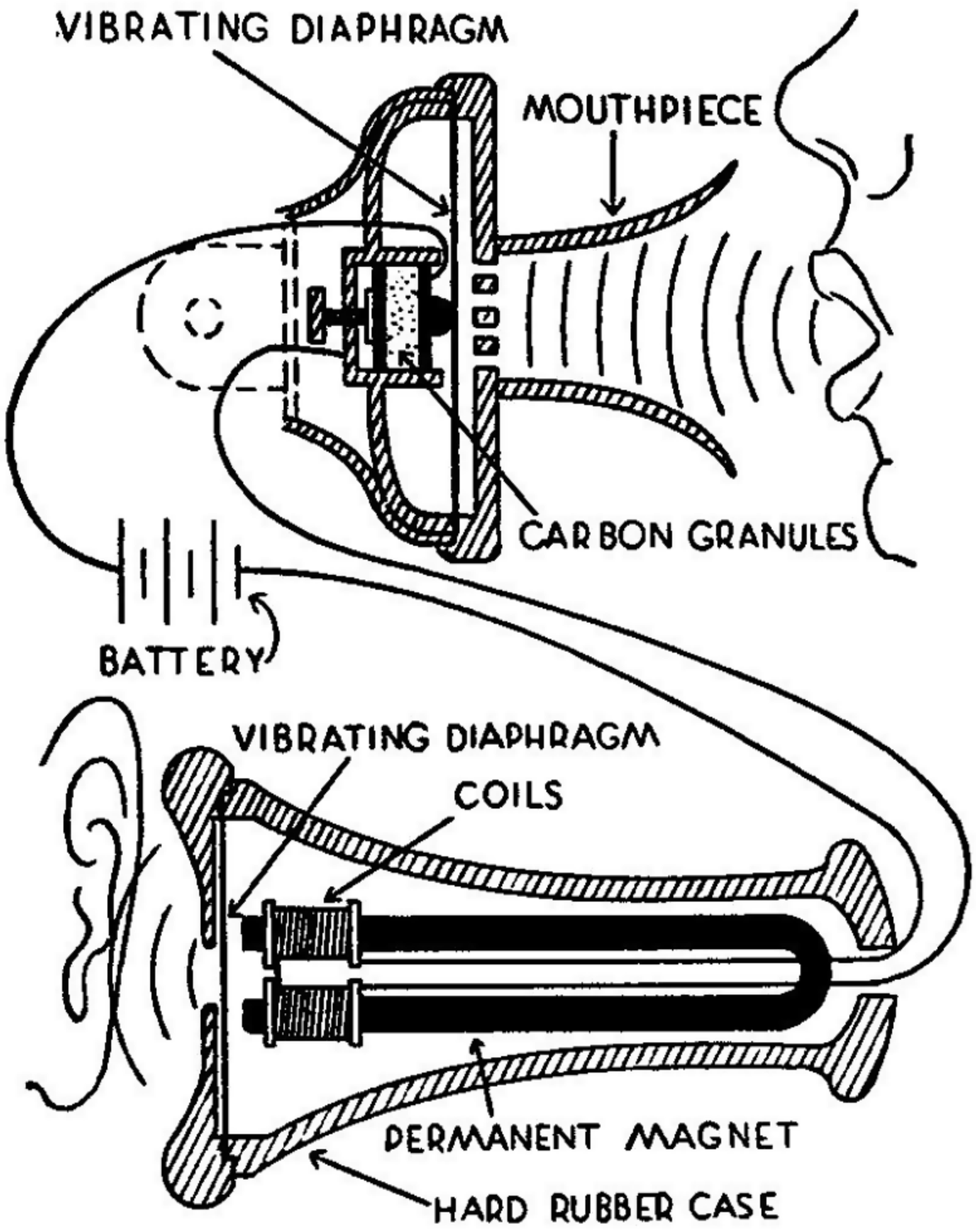


FIG. 217. Diagram of a simple telephone circuit.

of setting an iron diaphragm in vibration with sound waves and then transmitting the vibration electrically to another diaphragm. In Bell's first successful telephone, one with which he was able to talk only to the next room, the transmitter and receiver were exactly alike. Permanent magnets were used and the circuit contained no battery or other source of current. In Fig. 217 is shown a modern telephone circuit of the simplest construction. It embodies the essential features of all electrical telephonic communication. The receiver operates on the same principle as did Bell's. But the transmitter is quite different from the one developed by the earlier inventor. This one, which is called a carbon-granule transmitter, is the outgrowth of one built around 1877 by Thomas Edison. Let us trace the sequence of events

as they occur when speech is transmitted by means of this telephone circuit. First, it should be noted that there is a closed electric circuit, and that as the current flows in this circuit it passes through a small box of carbon granules in the transmitter, and through the coils of an electromagnet in the receiver. As the sound waves set up by the speaker's voice enter the transmitter, they strike against the *vibrating diaphragm*, making it vibrate with the same frequencies as those in the voice waves. As the diaphragm moves back and forth it communicates its motion to the button which is attached to one side of the box containing the carbon. As a result, the carbon granules are alternately packed more tightly together and then loosened up again as the diaphragm vibrates. Now, loosely packed bits of carbon offer greater resistance to the flow of electricity than do tightly packed ones. Therefore, the electric current in the circuit pulsates, that is, increases and decreases, at the same rate as the diaphragm vibrates. As this pulsating current flows through the coils of the electromagnet in the receiver, it causes the strength of the magnet to vary. Because of this varying magnetic strength, the *diaphragm* in the receiver will be set in vibration and will initiate a new set of condensations and rarefactions in the air, that is, will start out new sound waves. It is these that the listener hears. The new waves will be just like the original ones in frequency. So we see that it is not strictly correct to say that one's speech is transmitted by telephone. What does occur in any telephone system, no matter how complicated, is that the voice frequencies of the speaker set up varying electrical currents which have similar frequencies, which in turn put into vibration a diaphragm that produces a new set of sound waves that are essentially like the original ones.

Bell is credited with having invented the telephone, but a list of those whose contributions made possible its growth into the world-wide system of today would contain hundreds of names. We can mention here but a few of the most outstanding improvements made by these men. An important early one was the development of the switchboard, on which device alone one man filed more than 600 patents. Without it, or some equivalent device, one would have to have a separate pair of wires to the home or office of every person with whom he wished to converse. The modern automatic switchboard and dial telephone system has greatly increased the speed with which calls can

be handled. Another problem that made its appearance rather early in telephone history was what to do with the wires. In New York City, by 1887, some poles were as much as 90 feet in height and had 50 crossarms. One can imagine the result of trying to string the approximately 15,000,000 miles of telephone wire in New York today on poles. The underground cable proved to be the answer but it was perfected only after long research.

The development of the vacuum tube, so important in radio, also made possible the high quality of long distance telephone communication that we know today. By a combination of telephone and radio it is now possible to talk satisfactorily to points all over Europe, Asia, Africa, and South America, as well as to ships at sea and airplanes in flight. Although no telephone communication is possible over existing ocean cables, modern research indicates that today it would be possible to make a cable that could be used for speech. The necessities of modern radio have resulted in vast improvement in the quality of telephone transmission. As long as nothing but conversation was carried over the wires, the perfect transmission of a wide range of frequencies, or pitches, was not particularly important. But when it became desirable to carry by wire radio programs such as symphony orchestra concerts, accurate transmission of both high and low tones became important, and research solved the problem. Incidentally, the wonders of modern telephony are seldom appreciated by listening radio audiences. It is not uncommon to hear someone exclaim over the marvel of radio while listening to a network program originating perhaps 2,000 miles away and being broadcast by a local station. As a matter of fact the program is coming to the listener 1,980 or 1,990 miles by telephone and only the last 10 or 20 by radio.

The Radio

Some of the characteristics of radio waves were discussed in Chapter 31. It will be recalled that, as far as their nature is concerned, they differ from light waves only in that they are much longer, far too long to be detected by the human eye. In discussing them a little further here it is our purpose to show something of how they are generated and received. It is interesting to note that the great English mathematical physicist, James Clerk Maxwell (1831–

1879), predicted, on theoretical grounds, the existence of waves of this sort a number of years before they were actually discovered. It was in 1888 that a German physicist, Heinrich Hertz (1857–1899) demonstrated that there were such waves and that they traveled through space. His experimental set-up, shown in diagrammatic form in Fig. 218, consisted principally of two spark gaps. The small spheres of gap *A* were connected to a high voltage source of alternating current. The sides of spark gap *B* were connected by a wire slide which could be moved back and forth, thus changing the length of the conducting path from one sphere around to the other. Hertz found that when sparks were jumping across gap *A* as a result of the applied voltage, there was a certain location of the slide for which sparks would also jump across gap *B*, although there was no battery or similar source of electric current in the right-hand circuit. In other words, radio waves were emitted by *A*, and when these waves were cut by the other conductor, a current was set up in *B*. Although Hertz never recognized the communication possibilities of his experiment, what he really had in the above set-up was a very simple radio transmitter and receiver. It remained for Guglielmo Marconi (1874–1937) to see how these Hertzian waves might be used to transmit signals from one place to another. In the early days of radio, or wireless telegraphy as it was known then, a spark gap was employed in transmission. When a spark discharge occurred at the gap there was a surge of electrons back and forth between the ground and the antenna, as a result of which long electromagnetic waves were sent out through space. When these were cut by the aerial of the receiving set, electric currents were induced in the wires and signals were heard in the headphones. However, outfits of this sort could be used only for the transmission of code. Radio, as we know it today, was made possible by the development of the vacuum tube.

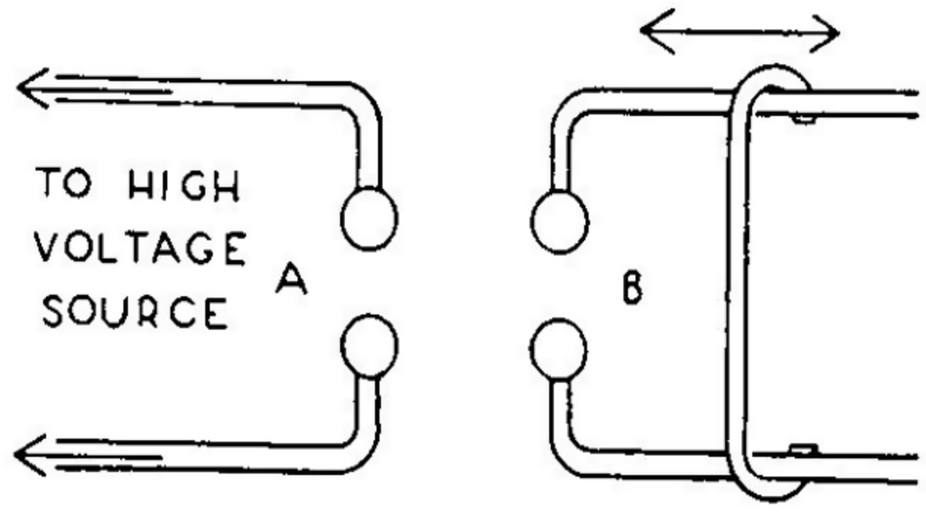


FIG. 218. Diagram of essential parts of Hertz device.

The various types of vacuum tubes were discussed in Chapter 27. Also at that time the operation of a simple radio circuit was described and so will not be repeated here.

It should be noted that the sequence of events occurring in radio communication involves one more stage than that which takes place in telephone transmission. Now, sound waves entering a microphone set up corresponding electrical variations in the transmitter as a result of which electromagnetic waves are sent out through space. These induce varying frequencies in the electrical circuit of the receiver which in turn cause the loudspeaker to emit a new set of sound waves essentially duplicating the original ones.

Other Communication Methods

There are a number of other electrical communication procedures which represent outgrowths and combinations of the three just discussed. We will mention several of these briefly.

The Printing Telegraph. — Several forms of this sort of machine have numerous applications today. One is the stock ticker which records automatically on a paper tape various information about stock sales. In this device the printing is done by a small wheel which has letters around its rim. It can be made to rotate in synchronization with a similar wheel in the transmitting machine. When the sending wheel is turned to a certain letter, the ticker wheel turns to the same one and prints it. A more recent development in the field of the recording telegraph is the teletypewriter. This machine resembles a very large ordinary typewriter in that it has a keyboard and a set of type bars, the former constituting the transmitter and the latter the receiver. Suppose it is being used to communicate between New York and Chicago. An operator in the former city simply types out the message. As each key is struck, a series of electrical circuits are closed as a result of which that letter is typed on the machine in the receiving city. Although the typing is being done in New York, the message appears on the paper in the Chicago machine. It is possible to connect machines in such a way that the material typed on one will be received by all the rest, just as telephone lines can be connected so that one person's voice can be heard on many phones. Such an application of the teletype is made by the United States Weather Bureau. A whole series of weather stations are connected in a sort of chain. In regular sequence a teletype operator at each one transmits the weather conditions in that locality and the data are recorded on all the other ma-

chines in the group. The device is also widely used today by other government bureaus as well as news services and business houses.

Pictures by Wire and Radio. — The photoelectric cell has made possible the transmission of photographs and other pictures by both wire and radio. It will be recalled from our previous discussion of this device that it emits electrons when light falls on it. It is also true that the magnitude of the current thus generated depends on the brightness of the light. Therefore, it provides a means of converting variations in brightness of light into variations in an electric current. Now a photograph is composed of areas of varying "lightness" and "darkness." In one type of sending device a beam of light passes through a negative of the picture to be sent and falls on a photoelectric cell. As the light beam moves back and forth over the film, the variations in its "darkness" are converted into electrical current pulsations. At the receiving end a light whose brightness is governed by the varying electric current passes in similar fashion over unexposed photographic paper reprinting the picture. The transmission of pictures by radio is similar except that there is the intermediate step of converting the electrical variations into radio waves and then back into electric currents again. In television, the principle of operation is essentially the same. Differences are due to the fact that now the entire scene must be transmitted once every sixteenth of a second whereas in ordinary wirephoto as much as 20 minutes may be required to transmit a single picture. To accomplish this enormous speeding up, a number of specialized and highly complicated devices have been developed.

Sound on a Beam of Light. — We have seen in connection with the telephone how it is possible to convert varying sound frequencies into electric ones, and in connection with wirephoto how a similar conversion between light of varying brightness and electrical current can be made. Consequently, one might expect to be able to use electrical currents as a sort of "middleman" and go from light to sound and vice versa. Such a procedure is exemplified in sound motion pictures. When the film is made, sound which enters the microphones sets up varying electric currents which in turn control a light that prints the sound track on the film as lines of varying density. Then, when the film is shown, a tiny light beam in the projector passes

through this sound track and falls on a photoelectric cell. The variations in density of the film cause the light that strikes the cell to vary in brightness which in turn produces pulsations in the electric current going through the cell. These are converted into sound in the loudspeakers much as the change is accomplished in a telephone receiver. Essentially the same conversion of light variations to sound frequencies occurs in the type of modern phonograph that uses a beam of light in reproducing sound from records.

Questions

Discussion

1. List as many different methods of communication as you can and rate each one as good or poor with regard to each of the four characteristics mentioned on page 622.

2. Look up some information about existing evidence with regard to man's earliest use of words.

3. Look up some information about the history of our alphabet.

4. Look up some information about an automatic type-setting machine called the Monotype and state how it differs from the Linotype.

5. Discuss similarities and differences between stereotyping and electrotyping.

6. Find out what the function of a "relay" is in a telegraph system.

7. Look up information concerning the various types of printing presses and find out which kind is employed in a local newspaper and what its capacity is.

8. Draw a diagram of a telephone receiver and explain the function of the electromagnet in it.

9. Draw a diagram of a carbon box transmitter and explain the function of the carbon granules.

10. How did Morse's first telegraph differ from the one devised by Wheatstone and Cooke?

11. Trace the energy transformations that occur in going from the individual who sends a message to the one who receives it, in the case of the telegraph, the telephone, the radio, the teletypewriter, and sound motion pictures.

12. Mention one way in which developments in radio have made the telephone a more efficient and valuable device.

13. Describe the experiment in which Heinrich Hertz first transmitted and received radio waves.

14. Find out something about the problems of color television as compared with those of black-and-white television.

Multiple Choice

1. Arrange the following in chronological sequence. (a) First use of automatic type-setting. (b) Earliest use of block printing. (c) Invention of the Linotype. (d) Invention of the teletypewriter. (e) Invention of movable type.

2. Which one of the following is the most important in the functioning of a simple telegraph system? (a) Three-electrode vacuum tube. (b) Two-electrode vacuum tube. (c) Magnetic needle. (d) Carbon granule box. (e) Electromagnet.

3. A container of carbon granules is used in one of the instruments discussed in this chapter in the transformation of energy from (a) electrical to light, (b) sound to electrical, (c) magnetic to sound, (d) sound to light, (e) electrical to sound.

4. The first important application of radio waves to practical communication was made by (a) Hertz, (b) Marconi, (c) Morse, (d) Mergenthaler, (e) Wheatstone.

5. In which of the following communication methods does the energy sequence "sound to electrical to electromagnetic to electrical to sound" occur? (a) Television. (b) Radio. (c) Sound motion pictures. (d) Teletype. (e) Telephone.

True-False

1. It is believed that verbs entered the spoken language of man earlier than nouns.

2. Man's earliest use of writing was for the purpose of preparing messages for runners to carry between tribes.

3. The first use of block printing in Europe occurred around 1450 A.D.

4. The names of Coster and Gutenberg are associated with the development of movable type.

5. Curved printing plates are employed in the rotary type of printing press.

6. The modern teletypewriter is more closely related to the radio than it is to the telegraph or telephone.

7. Thomas Edison had more to do with the development of the modern telephone receiver than he did with the transmitter.

8. The photoelectric cell plays an important part in the operation of a sound motion picture projector.

9. All telephonic communication carried on today between North America and Europe makes use of radio.

10. Indian smoke signal communication satisfies three of the four tests given at the beginning of this chapter as criteria of a good communication system.

Chapter 33

ILLUMINATION

VISIBLE light has already been defined as the rather narrow range of wave lengths in the electromagnetic spectrum to which the eye is sensitive. For many centuries the sun was man's only source of light and it still is his most important one. But there came a time when he began to use artificial light and, as civilization advanced, he found increasing need for such illumination. It is this question of artificial light that will concern us in this chapter. First, we shall consider some of the general problems of illumination, and then discuss the various types of artificial illuminants that man has used in the past and is using today.

GENERAL ILLUMINATION PROBLEMS

There are certain general points to be considered with regard to any source of artificial light. We shall take up briefly four of these.

Intensity

The intensity of illumination on any surface or object is governed by two factors. These are the actual intensity of the source of light and the distance of the illuminated surface from that source.

Intensity of the Source. — At the time when man first became interested in measuring and comparing light intensities, the candle was a common source of light. Therefore, it was natural that a unit called the candle-power should become the basis for intensity measurements. The candle-power was defined simply as the intensity of a so-called standard candle. Since candles varied greatly in brightness it was necessary to specify carefully just what kind of candle should be called standard. The detailed nature of these specifications is shown by the fact that they included the kind of wax and its chemical treatment; kind of thread used in the wick, as well as the number of threads, their size and degree of twist; height, thickness, and weight of finished prod-

uct; the height of the flame; and the temperature, pressure, and humidity of the atmosphere in which it burned. Even with all these items defined, standard candles varied considerably in intensity. When liquid light sources came into use it was found that their intensities could be more accurately controlled and so, although the name candle-power was retained, the definition of the unit was stated in terms of the newer type of source. Today the so-called international standard candle is defined in terms of a certain type of lamp which burns pentane vapor in air under specified conditions. Such lamps are used in standardizing laboratories. Most other laboratories use as standards, incandescent electric bulbs which have been tested by the Bureau of Standards and rated as being a certain candle-power when operated under particular conditions of current and voltage. Until comparatively recently all light bulbs were rated in candle-power and so labeled. But it happens that the intensity of an ordinary incandescent type of electric lamp is approximately proportional to the rate at which it consumes energy, that is, to its power. So the practice today is to rate bulbs in terms of the latter quantity using the watt as a unit, except for automobile headlight bulbs which are still rated in terms of candle-power. A 50-watt bulb has about twice the intensity of a 25-watt one, a 100-watt lamp about twice that of the 50-watt and so forth. Ordinary incandescent bulbs of medium wattages have luminous intensities amounting to approximately 1 candle-power per watt. It should be pointed out that intensities are not proportional to power consumption if one compares incandescent lamps with some of the newer types of electric illuminants. A 40-watt fluorescent tube, for instance, gives between two and three times as much light as a 40-watt incandescent bulb.

Relation of Intensity to Distance. — It is obvious that moving closer to the light increases the intensity of illumination on the book or magazine that one is reading. Exactly how much the intensity increases for a given change in distance is not quite so obvious. It happens that the law which expresses the numerical relationship between these two quantities is a consequence of an experimental fact discussed back in Chapter 30, namely, that for all ordinary purposes light can be thought of as traveling in straight lines.

In Fig. 219 there is shown a source of light S. Although it is emitting light in all directions, we are interested here in that which is

directed down toward the three plates *A*, *B*, and *C*, located at distances of 1, 2, and 3 feet respectively from S. First, we shall define a unit of illumination. If S is a 1-candle-power source of light then the intensity of illumination on *A* is said to be 1 *foot-candle*. In other words, the unit called the foot-candle may be defined as the illumination at a point 1 foot from a light source of 1 candle-power. Now let us inquire about the intensity of illumination on plate *B*. The dotted lines are drawn from S past the corners of *A* and so represent rays which define the limits of the light which strikes the top plate. *B* then is the square that can be drawn within these light ray limits when a point 2 feet below S is reached. By plane geometry it can be shown that the area of *B* is four times that of *A*. Now, it will be noted that if *A* is removed, the total amount of light that falls on *B* is exactly the same as that which originally fell on *A*. Thus, we have a situation in which the same amount of light is " spread out " over an area four times as large.

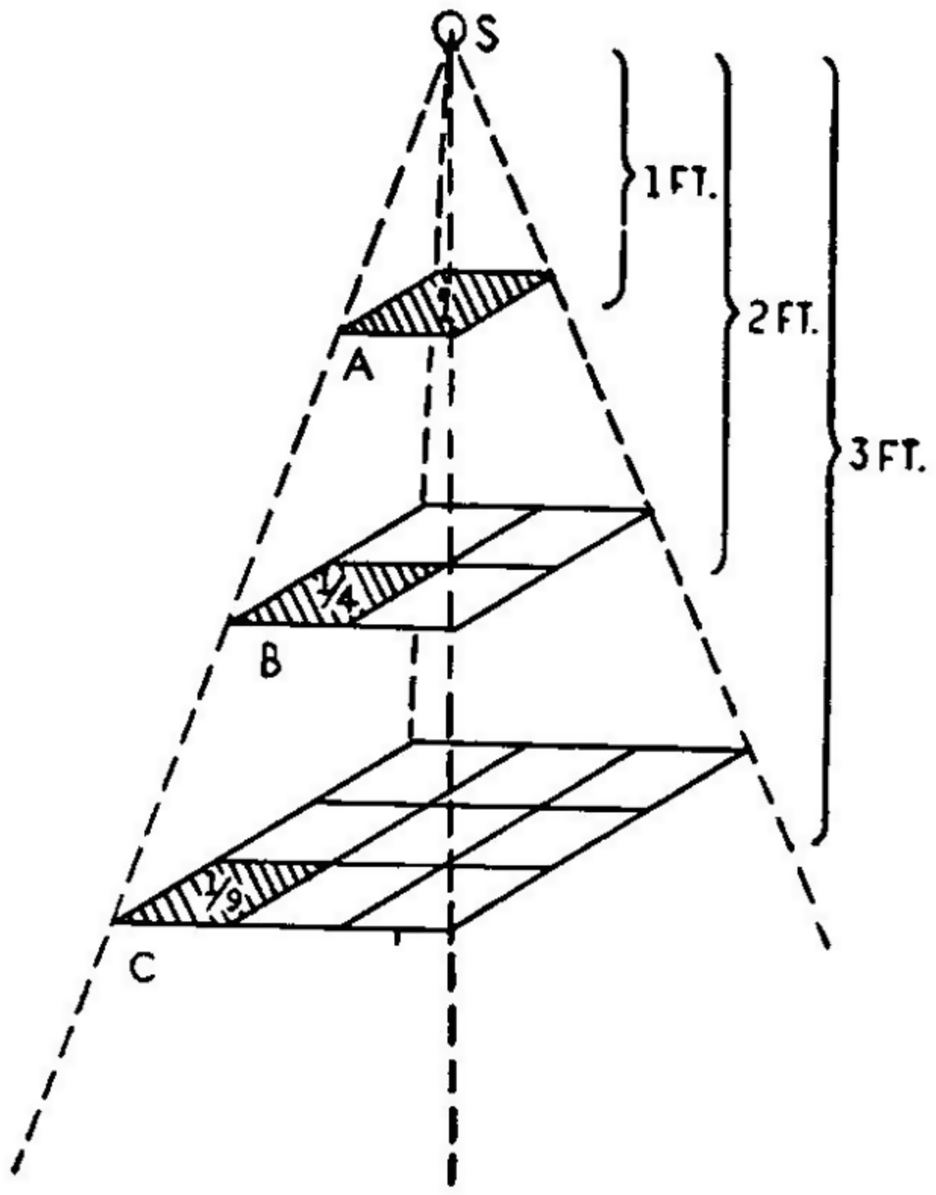


FIG. 219. Diagram illustrating the manner in which light intensity varies with the distance from the source.

If we were discussing water instead of light, it would be obvious that the amount of water which had been 1 inch deep on a surface of 1 square foot would only be a quarter of an inch deep if poured over a surface of 4 square feet. Similarly here, the intensity of illumination on *B* is only $\frac{1}{4}$ of what it was on *A*. If we go on down to a point 3 feet below S and draw in the square *C* as defined by the dotted lines, we find it to have 9 times the area of *A*. Consequently, its illumination is $\frac{1}{9}$ that of the top square. Thus we find that,

> *doubling the distance from the source makes the illumination* $\frac{1}{2}^2$ *or* $\frac{1}{4}$ *as great; tripling the distance from the source makes the illumination* $\frac{1}{3}^2$ *or* $\frac{1}{9}$ *as great, and so forth.*

In general terms then, the intensity of illumination can be said to vary inversely with the square of the distance from the source. Other ex-

amples of inverse square laws that have been mentioned previously in this book include the law of gravitation and those of electric and magnetic attraction.

Up until quite recently almost all reading and work done at home and in industry under artificial light was done at intensities which were far too low for comfort and for the good of human eyes. Such illumination was commonly as little as 2 or 3 foot-candles. Experiments performed in recent years indicate that while necessary intensities vary considerably with the task being performed, illuminations of several hundred foot-candles give the best vision and a minimum of eyestrain for fairly difficult visual work. That even several hundred foot-candles do not represent an extreme degree of brightness may be seen from the fact that outdoor illuminations may range from 1,000 foot-candles in the shade to as much as 9,000 or 10,000 in direct sunlight. Table 34 gives what Dr. Matthew Luckiesh, director of lighting research laboratories of General Electric Company, calls conservative foot-candle recommendations for various visual tasks.

TABLE 34.[1] CONSERVATIVE FOOT-CANDLE RECOMMENDATIONS ON A NATIONAL BASIS OF CHARACTERISTICS OF THE VISUAL TASK AND REQUIREMENTS OF PERFORMANCE

100 Foot-candles or More. — For very severe and prolonged tasks such as fine needlework, fine engraving, fine penwork, fine assembly, sewing on dark goods and discrimination of fine details of low contrast, as in inspection.

50 to 100 Foot-candles. — For severe and prolonged tasks, such as proof-reading, drafting, difficult reading, watch repairing, fine machine-work, average sewing and other needlework.

20 to 50 Foot-candles. — For moderately critical and prolonged tasks, such as clerical work, ordinary reading, common benchwork and average sewing and other needlework on light goods.

10 to 20 Foot-candles. — For moderate and prolonged tasks of office and factory and when not prolonged, ordinary reading and sewing on light goods.

5 to 10 Foot-candles. — For visually controlled work in which seeing is important, but more or less interrupted or casual and does not involve discrimination of fine details or low contrasts.

0 to 5 Foot-candles. — The danger zone for severe visual tasks, and for quick and certain seeing. Satisfactory for perceiving larger objects and for casual seeing.

[1] Luckiesh and Moss, *The Science of Seeing*, p. 345, Table LVIII. New York: D. Van Nostrand Co., 1937.

Shadows and Glare

It has been found that the intensity of the illumination in a room should be fairly uniform if eyestrain is to be avoided. In as much as one common reason for non-uniformity is the existence of sharp shadows, the causes of such shadows should be eliminated as far as possible. The frosted type of electric light bulb produces less pronounced shadows than did the old-fashioned clear type, since in the former the light comes from the entire surface instead of from a small, very bright filament. Still better, from the standpoint of eye comfort, is indirect lighting, in which much of the illumination of a room results from diffuse reflection of light by the ceiling and walls. One of the advantages of the new fluorescent lighting is the fact that the comparatively large area from which the light comes tends to prevent the formation of shadows.

Glare represents one of the most common defects in lighting as well as one of the most serious from the standpoint of eyestrain. It may result from any of several causes. One is simply excessive brightness of the light source, in which case the remedy is obvious. Another is too much contrast, such as occurs when one uses a bright reading lamp in an otherwise dark or very dimly lighted room. Here the remedy is to increase the illumination in the rest of the room. One of the most common causes of glare is reflection from surfaces such as glass, polished wood, shiny paper, wet pavements, or snow. Remedies include the use of dark glasses, and variation in angle or direction at which the light strikes the surface. The applications of Polaroid materials in this connection have already been mentioned in Chapter 30 in the discussion of polarization. As was pointed out at that time, they seem to have particularly valuable possibilities for the elimination of the glare of automobile headlights, which constitutes one of the major lighting problems in the United States. Another approach to the solution of this problem which is just beginning to be applied is the flood-lighting of highways so that automobile headlights will not need to be so bright.

Color Distribution

It was pointed out in Chapter 31 that the color which an object appears depends to some extent on the assortment of shades

that happen to be present in the light that is being used to illuminate it. Since so much of our seeing is done by daylight we think of it as standard and call the appearance of any material in daylight its " natural " color. So for most purposes it is desirable that an artificial light source approximate daylight as nearly as possible. Anyone who has looked at objects or at another person under a sodium highway light or a mercury vapor lamp knows that either of these produces an extremely unnatural appearance. Therefore, such light sources are not likely to become popular for ordinary home lighting. The common tungsten filament incandescent light approximates daylight but is lacking to some extent in the blue and violet part of the spectrum. That it isn't exactly the same as daylight is evidenced by the fact that articles of clothing purchased under artificial light frequently appear a somewhat different shade when one views them out of doors. By combining mercury light, which has an excess of the blue-violet wave lengths, with that from incandescent tungsten filaments a close approach to daylight can be obtained.

On the other hand, there are some purposes for which it is desirable to have light of as nearly one color as possible instead of a combination of all colors as in white light. For one thing, there is evidence to indicate that objects can be seen with a greater degree of sharpness in one-color light, such as is emitted by sodium vapor lamps, than in white light. Certain color-matching operations can also be carried out more easily under illumination of a single color.

Cost

The cost of obtaining a given amount of artificial light has decreased so enormously since 1800 that a few figures on the subject may be of interest. Two factors have contributed to this decrease. For one thing science has been able to produce increasingly better sources of light. The other reason has been the progressive lowering in the cost of fuel used in the generation of light. The comparative cost of light at the beginning and end of the nineteenth century has been stated by Dr. Luckiesh in the following terms.[1] In 1800 a family could burn 5 candles for 3 hours an evening at a total annual cost of $20. By 1900, using new sources, the same amount of money would buy a

[1] Luckiesh, M., *Artificial Light.* New York: Century Co., 1920.

quantity of light the equivalent of that emitted by 154 candles burned five hours a night for a year. This means that the number of candle-hours of light that could be obtained for a given price increased in one century more than fifty fold, where a candle-hour is the light given off by a standard candle in one hour. Table 35 shows similar facts stated somewhat differently.

TABLE 35.[1] TABLE SHOWING THE APPROXIMATE TOTAL COST OF 1,000 CANDLE-HOURS FOR VARIOUS PERIODS

Period	*Source*	*Cost of* 1,000 *Candle-Hours*
1800 to 1850	Sperm oil	$2.40
	Tallow candle	5.00
1850 to 1865	Kerosene	1.65
	Tallow candle	6.85
1865 to 1875	Kerosene	.75
	Tallow candle	6.25
	Gas, open flame	.90
1875 to 1885	Kerosene	.25
	Gas, open flame	.60
1885 to 1895	Kerosene	.15
	Gas, open flame	.40
1895 to 1915	Gas mantle	.07
	Carbon filament	.38
	Metallized filament	.28
	Tungsten filament (vacuum)	.12
	Tungsten filament (gas-filled)	.07

[1] Luckiesh, M., *op. cit.*, p. 218.

Assuming an electrical rate of 5 cents per kilowatt-hour, modern incandescent bulbs furnish 1,000 candle-hours of illumination at a cost of from 4 to 6 cents, depending on the particular size and type of lamp. With fluorescent lighting the price of the same amount of light is in the neighborhood of 2 cents.

Sources of Artificial Light

There are four general methods by which the one octave of electromagnetic radiation that we call visible light is generated. With respect to each of these we shall consider briefly its nature, its present practical applications if any, and its future possibilities in the light of present knowledge.

Incandescence

Let us assume that we have prepared an oven in which a piece of iron, a brick, or other object can be heated to a high temperature. Performing the experiment, we note the following sequence of events. As soon as the object becomes a little warmer than its surroundings it begins to emit radiation that we can feel, that is, heat radiation. This radiation consists of the " longer than red " waves that, in Chapter 31, we called infra-red rays. At a temperature of approximately 900° F., the first visible radiation is given off, when the object becomes dull red in color. By about 2,200° F. it is white hot and is said to be *incandescent*. So an incandescent light source can be defined as one which emits light because of its high temperature. It is important to note that the heat energy such a source emits is always far greater than that which is given off in the form of light, which is the reason for the relatively low efficiency characteristic of all incandescent sources.

Pre-electric Artificial Light. — With one exception, all of man's sources of light until very recent times, including the most commonly used electrical ones, have been of the incandescent type. The exception was the practice of some primitive peoples of imprisoning fireflies and glow-worms in receptacles of various sorts and using them as light sources. As might be expected, such illuminants did not prove particularly satisfactory and never attained any great importance. The principal sources of light that man has used since prehistoric times, together with some information about each, are listed chronologically in Table 36. A few additional comments should be made about the early devices which man developed to provide himself with light at times when, and in places where, daylight was not available.

For one thing, the importance of carbon in the field of illumination should be noted. In every one of the sources of artificial light

introduced before the advent of electricity, and in several since that time, carbon is the material which becomes incandescent and emits light. In torches, the light is furnished by the combustion of wood or other material. In the various types of oil lamps and gas lighting fixtures, hydrocarbons in the fuel burn and generate the necessary heat. The high temperatures thus produced cause some free carbon to be formed from the hydrocarbons, the floating particles of which become incandescent. If the temperature is not sufficiently high to raise all the particles to incandescence many of them escape as smoke or soot.

TABLE 36. A SUMMARY OF THE PRINCIPAL INCANDESCENT LIGHT SOURCES
(ADAPTED FROM A TABLE PREPARED BY PROF. F. F. HOUSEHOLDER, UNIVERSITY OF AKRON)

Approximate Date	*Description*	*Intrinsic Brilliancy*	*Color*	*Remarks*
Prehistoric	Torch	1.2 c.p. per sq. in.	Yellowish-orange	Earliest known artificial source. Smoky, unsteady, unclean, but convenient.
Early Chinese, Greek, and Roman civilizations	Oil lamp, using sperm and olive oil, fats, etc.	3–8 c.p. per sq. in.	Orange	First step toward modern oil lamps. Big improvement over torches. Smoky and unsteady.
200 A.D.	Candle	3–8 c.p. per sq. in.	Orange	Portable and fairly safe. Clean, odorless, convenient. Small units only. Became an intensity standard.
1779	Bare gas flame	3–8 c.p. per sq. in.	Orange to yellow	Fairly steady, odorless, convenient. Small units. Not portable. Some danger of explosion or asphyxiation.

Approximate Date	*Description*	*Intrinsic Brilliancy*	*Color*	*Remarks*
1853	Kerosene lamp	4–9 c.p. per sq. in.	Orange to yellow	Cheap, clean, portable, fairly safe. Outdoor use if well protected. Requires good air supply.
1870	Carbon arc	10,000 to 100,000 c.p. per sq. in.	Bluish-white to violet	Now almost obsolete in old form. Large units only. Efficiency far below metal filament lamps. Not portable.
1879	Carbon filament	300–500 c.p. per sq. in.	Yellowish-white	Clean, safe, convenient. Now replaced by metal filament lamps and other newer sources.
1886	Gas (Wellsbach mantle)	20–50 c.p. per sq. in.	Greenish to amber hue	Clean, odorless. Fairly steady. Small units. Some danger of asphyxiation.
1890	Acetylene gas	75–100 c.p. per sq. in.	Yellowish tint	Nearly 15 times the illuminating power of natural gas. Dangerous because the gas is rather unstable.
892	Mercury arc, glass enclosed	10–15 c.p. per sq. in.	Greenish-blue	Large units only. Good for seeing details in black and white. Not good where color distinction is required.

Approximate Date	*Description*	*Intrinsic Brilliancy*	*Color*	*Remarks*
1894	Moore tube	.67 c.p. per sq. in.	Varied with gas; almost daylight with CO_2	Forerunner of modern neon type tube. Large units only. Used various gases, chiefly CO_2 and N. Trouble with gas disappearing.
1899	Flaming electric arc	5,000 c.p. per sq. in.	Yellowish tint	Large units only. Outdoor lighting. Constant sputtering.
1904	Tantalum filament	750 c.p. per sq. in.	Yellowish tint	Short life. Easily broken. Convenient, clean, semi-portable. Medium sized units.
1906	Tungsten Mazda "B"	1,060 c.p. per sq. in.	Daylight	Mazda "B" lamps are vacuum filled. Obtainable in clear glass or inside colored in various hues. Can be used out of doors without protection.
1909	Mercury arc, quartz enclosed	800 c.p. per sq. in.	Bluish-white	Large units only. Rich in ultra-violet. Fairly long life. More red than in glass enclosed type.

Approximate Date	*Description*	*Intrinsic Brilliancy*	*Color*	*Remarks*
1913	Tungsten Mazda "C"	3,000 c.p. for 50-watt to 8,400 c.p. per sq. in. for 2,000-watt	Nearly white	Mazda C lamps are gas-filled. Convenient, clean, safe. All sized units. Adaptable for almost all lighting requirements. If used out of doors should be protected from rain and snow.
1922	Neon light		Deep orange-red	Used on high voltage alternating current only. High visibility. Chief application in advertising.
1930	Sodium vapor		Orange-yellow	High efficiency. Monochromatic. Tube contains neon or argon to carry arc until sodium is vaporized. Used for street and outdoor lighting.
1939	Fluorescent tube		Wide range of possible colors	Ultra-violet light from mercury vapor strikes fluorescent materials coated on inside of tube. Two or three times as efficient as filament lamps.

When bare-flame gas lighting was first introduced it proved to be a big improvement over candles and oil lamps in convenience, safety, and cost of operation. However, the gas mantle, which did not appear until after the electric lamp had been invented, was responsible for the very great importance which gas lighting attained before it was finally largely superseded by electricity.

Incandescent Electric Lighting. — The first electrical illuminating device to be developed was the arc light. Although it did not become commercially important until about 1870, such a light was demonstrated as early as 1808 by Sir Humphry Davy. At that time a subscription was taken among the members of the Royal Society to provide funds for building an electric battery of 2,000 voltaic cells. This battery was used to furnish the current for a lecture demonstration by Davy of an electric arc between charcoal rods. However, the practical use of such devices as sources of light had to await the development of generators that would provide large currents at low cost. Carbon rods are commonly used today in arc lights. Here again the substance which becomes incandescent is chiefly carbon, although a core of some other material is frequently put in the rod to add particular properties to the light that is emitted. Arc sources give off a very intense light and formerly found their principal application in street lights. But today they have been replaced for this purpose almost entirely by large-unit filament lamps. Arc lamps still find considerable use in large motion picture projection outfits and in various forms of searchlights.

The first successful incandescent filament electric lamp was made by Thomas Edison in 1879. In it he used a carbon filament which he prepared by carbonizing a strip of paper. This was enclosed in a glass vessel which was then evacuated. Platinum wires sealed into the glass and attached to the filament served as lead-in wires. It was necessary to use platinum for this purpose because it was the only good conductor known at that time which would expand and contract with temperature changes at about the same rate as glass. Today platinum has been replaced by a copper-plated nickel-iron alloy which is much cheaper. It should be noted that we are still dealing with a source in which the incandescent material is carbon.

In the years following the introduction of the carbon filament

lamp Edison and others carried out extensive research looking toward its improvement. These experiments were along two general lines, namely, the development of better methods of making carbon filaments and the search for substances that might be used instead of carbon. The next important step forward came in 1906 when the tungsten filament lamp was first put on the market. Tungsten, which was discovered by Scheele in 1870, is one of the heaviest elements, has a fairly high electrical resistance, and a melting point above 5,000° F. The tungsten filaments used in early lamps were extremely fragile and it took many years of research before a method was developed for making ductile tungsten of high tensile strength which could be drawn out into very fine wires. Tungsten oxidizes rapidly at high temperatures and so a tungsten filament lamp that contained air would burn out as soon as the current was turned on. To prevent this, light bulbs, for a number of years, were evacuated as highly as possible. It has been found, however, that less blackening occurs and the lamp lasts longer if a gas of low chemical activity is put in it. Nitrogen was formerly used for this purpose but, as was mentioned in Chapter 15, it has almost entirely been superseded today by argon. Other improvements of recent years include inside frosting of bulbs, appreciable lengthening of life, and increase in efficiency, modern tungsten lamps being about six times as efficient as the old carbon filament type.

Two special types of lamps of particular interest to photographers have been developed in recent years. These are photoflash and photoflood bulbs. The former contain a little aluminum foil and oxygen at a pressure of approximately a fourth of an atmosphere. A filament serves to ignite the foil which produces a flash of light lasting a few hundredths of a second and having an intensity of several million candle-power. The photoflood lamp contains a tungsten filament which is designed for a voltage about two-thirds that of the usual household circuit. Thus, when used on a 110-volt circuit, it is overloaded and produces an intense light but burns out after approximately two hours of service.

Incandescent Efficiency. — A brief reference was made previously to the relatively low efficiency of incandescent light sources. It is important to note that this is inherent in the manner in which the light is produced. In other words, not much can be done about it. In any

incandescent light some substance has to be heated to a high temperature in order for it to give off light, and upwards of 95 per cent of the energy that it takes to produce this heating is given off in the form of radiation to which the human eye is not sensitive. Consequently, the over-all light efficiency of the best incandescent electric lamps is not greater than 5 or 6 per cent. This simply means that of every dollar's worth of electricity burned by the ordinary electric lights in your home, you get back 5 cents' worth or less in the form of light. Increasing the temperature to which the filament is raised tends to increase efficiency, but the possibilities of improvement by this means are limited by the melting point of the tungsten. Since just now various installations of fluorescent lighting are being made in homes, it should be mentioned that the above statement does not apply to them. Their characteristics are discussed in the next part of this chapter.

Electroluminescence

We are using the term "electroluminescence" to apply to light sources which are essentially the modern counterparts of the Geissler and Crookes cathode-ray tubes described in Chapter 26. In other words, these sources emit light as a result of the discharge of electricity between two electrodes in a partially evacuated glass tube. It will be recalled that the color of the light in such a case depends on the kind of gas that the tube contains.

Neon Lamps. — The neon tube was described in some detail in the reference just mentioned. The gas is at a pressure of something like half a centimeter of mercury and a voltage of around 15,000 volts is applied across the electrodes. Everyone is familiar with the orange-red light which such a source emits. Other colors may be obtained by using different gases or mixtures of gases and colored glass. Hydrogen glows a pale blue and helium, yellow. A brighter blue is obtained by mixing neon, argon, and mercury vapor together. The same mixture, put in an amber glass tube, gives green light. The efficiency of such sources is roughly twice that of incandescent lights. However, the first cost of the equipment is high. The light is not steady because of the fact that it goes out for an instant every time the alternating current reverses. One can demonstrate this by moving his hand back and forth between his eyes and a neon sign. Instead of moving

smoothly the fingers appear to go by a series of jumps. If the sign happens to be using the common 60-cycle alternating current, these jumps are one-hundred-twentieth of a second apart. This flicker can be partially eliminated by using two tubes, side by side, which are operated by currents that are out of phase or "out of step" with each other. No combination of gases has been developed that gives a good white light, which fact combined with the flicker mentioned above makes neon-type lamps unsuitable for ordinary household lighting purposes.

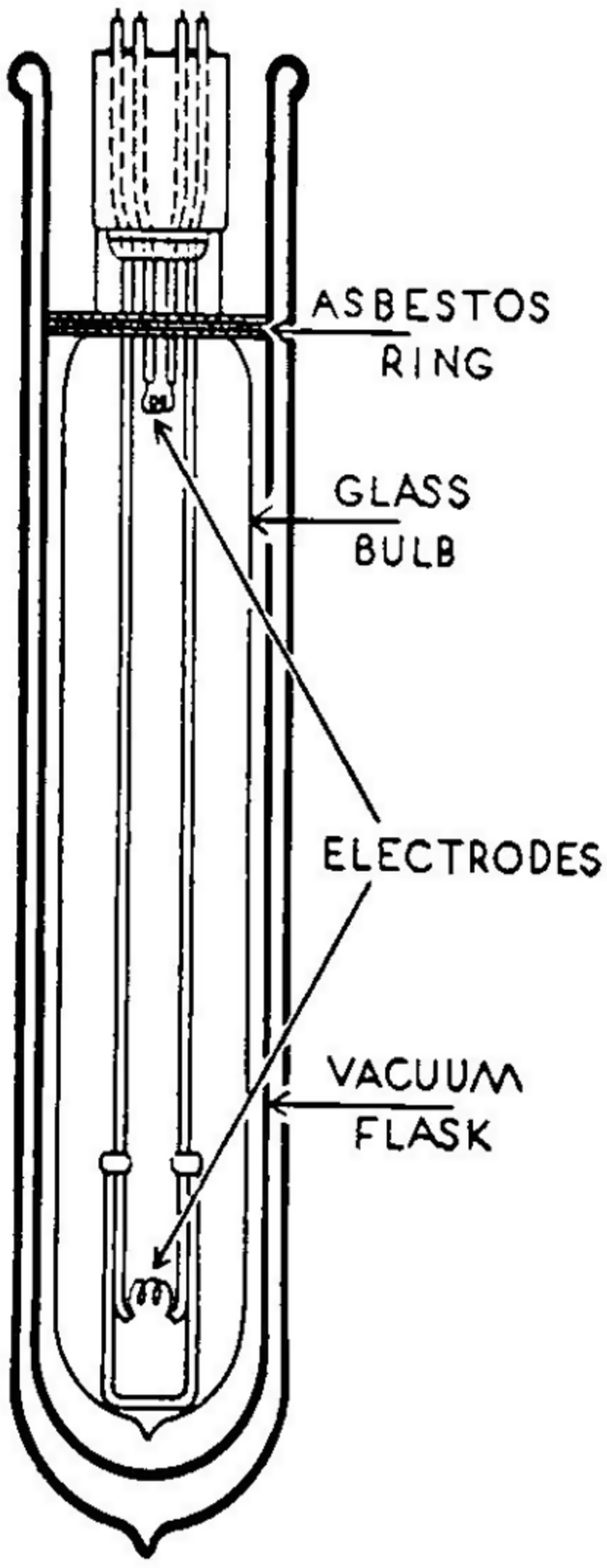


FIG. 220. Cross-section diagram of a sodium highway light bulb.

Sodium Lamps. — Although the sodium lamp operates much as does the neon, the tube must be constructed somewhat differently since at normal temperatures sodium is a solid. Such a tube is shown in Fig. 220. Each electrode is a small coil of tungsten wire. When the current is turned on these coils become incandescent. The heat thus generated serves to vaporize a piece of sodium, after which the device operates as a vapor lamp. A little neon is also placed in the tube to aid in the starting process. It is because of this that sodium highway lights give off a red light when first turned on, which later changes to yellow. They are from two to three times as efficient as incandescent lights and have found their principal application in highway lighting.

Mercury-Vapor Lamps. — Here also the operation is similar to that of a neon light although some means must be provided to vaporize the mercury to begin with. When the light is turned on a tungsten filament becomes hot and vaporizes some mercury in the bulb. However, the tungsten has a higher resistance than does the mercury vapor, so once the latter is formed a mercury arc is set up between the electrodes, and the filament ceases to carry much current. The light from these lamps is strong in green and blue but contains very little red. There is also a considerable amount of ultra-violet radiation in it. Mercury-vapor lamps made of quartz, which transmits ultra-violet, are

employed as sources of this very short-wave-length radiation for sterilizing purposes and other uses.

Fluorescent Lights. — Fluorescent lighting, which was first presented to the world at the 1939 New York and San Francisco fairs, marked an important step forward in the science of illumination. The applications found for it have increased so rapidly since its introduction that some 25,000,000 tubes were made in 1941, or about three times as many as in 1940. The lamp proper consists of a glass tube which contains some mercury and has an electrode sealed in each end. When it is operating the electric current flows through the mercury vapor just as it does in the mercury lamp described above except that here the current, voltage, and vapor pressure are so adjusted that the radiation is almost entirely ultra-violet in nature. The inside of the glass tube is coated with a powder which " fluoresces " in ultra-violet light. Now any fluorescent substance is simply one that has the ability to absorb rays of one wave length and then emit them again as longer waves. In this case the absorbed invisible ultra-violet rays are re-radiated as visible light. Different kinds of powder give different colors. Among those commonly used are the tungstates of magnesium and calcium and the silicates of zinc and cadmium.

The outstanding desirable feature of fluorescent lighting is its high efficiency, which is several times that of incandescent lighting. It is also true that the tubes last from two to three times as long as do the ordinary light bulbs. By the procedure of absorption and re-radiation the unsteadiness mentioned in connection with neon is eliminated. It also is possible to produce light of a great variety of shades and colors including a close approach to daylight. Another advantage, particularly in store lighting, is the almost complete elimination of shadows because of the relatively large surface from which the light comes. The first cost of fluorescent installations is somewhat higher than for the incandescent type.

There are a number of applications and possibilities of fluorescence that are quite interesting and novel, many of them not directly in the field of illumination. Museums have found it possible to use it to reproduce accurately the light from the northern sky so that pictures can be viewed under illumination similar to that in which they were painted. The fact that the powder does not have to be in the

tube but will fluoresce even when a considerable distance from the ultra-violet source has given rise to a number of uses. For identification purposes an invisible fluorescent paint can be used which becomes visible only when illuminated with ultra-violet rays. This has been applied to the marking of clothes in several laundries, and in at least one hospital for marking babies to avoid any chance of mixing them up. Subway and bomb shelter entrances in London are marked with fluorescent paint which gives off a glow when ultra-violet lamps are directed on it. Its possibilities for outdoor advertising at night are more or less obvious. Fluorescent tubes are available today in lengths all the way from six inches up to five feet.

Chemiluminescence and Autoluminescence

These last two types of light sources are mentioned simply to make the story complete since they have no important practical applications in the field of illumination today and seem to offer little in the way of future possibilities.

The most familiar example of a chemiluminescent source is the firefly. The evidence indicates that two chemical substances are involved in the production of light by the firefly. Their exact composition has not been determined but they have a protein-like nature. Neither one is phosphorus as has been commonly believed in the past. Light is emitted when one of these substances comes into contact with oxygen in the presence of the other which acts like a catalyst. It is worth noting that the firefly's light is some 95 per cent efficient which is enormously higher than the other sources we have discussed. Among fireflies their lights serve, not as illuminants, but as a means of bringing the sexes together. Incidentally, one writer [1] reports that in South America there is a species of firefly known locally as the automobile bug because it has a white light in front and a red one behind. Other forms of life which emit light because of chemical reactions include some types of bacteria, many mushrooms and other fungi, certain deep-sea fishes, numerous kinds of jellyfish, sponges, and many others.

A number of chemicals have been discovered which can be made to emit light in the laboratory quite apart from the life processes of

[1] Cottman, E. W., "Cold Light," *The Scientific American*, 6, 160, June, 1939.

any animal or plant. One of the most striking of these is a substance known commonly as " luminol " rather than by its complete name of 3-aminophthalhydrazide. It glows with a brilliant blue light upon oxidation, producing almost no heat.

Autoluminescence includes the light produced as a result of radioactive disintegration. So-called radium-coated watch dials and hands represent one application of this sort of illumination. The light is a result of the bombardment of certain materials by the rays emitted by the radioactive elements present. Any use of such light source on a large scale would be enormously expensive and in addition would be a definite health hazard.

Thus we may say that for some time to come most of man's illumination needs will be taken care of by incandescent and fluorescent sources. The former, although of low efficiency, provide satisfactory amounts of light at comparatively low cost. Although the latter is still too new for its future possibilities to be completely evaluated, they seem to be very promising in the light of its progress during the few years of its existence.

Questions

Discussion

1. Mention two factors that affect the intensity of illumination on any given surface.

2. Name and discuss three different light sources that have at some time or other in history been used as standards of illumination intensity.

3. In what sort of light sources may intensity be assumed to be approximately proportional to wattage?

4. Find as many laws as you can, mentioned earlier in this book, that can be described as inverse square laws.

5. Discuss methods of eliminating glare and shadows in home lighting.

6. Fig. 195 on page 580 may be assumed to represent the manner in which light is reflected from the curved reflector in an automobile headlight. Draw ray diagrams to show the result of placing the source of light either a little above or below the point indicated in the figure.

7. Find out how the deflected beam actually is obtained in automobile headlights and compare this with your drawings made in answer to the preceding question.

8. Assuming an average intensity of one candle-power per watt and using your local electric rates, calculate the cost of 1,000 candle-hours of light. Compare the answer with that given in Table 35.

9. Explain the meaning of the term "incandescence" and name five incandescent light sources.

10. Name three sources of light that are not incandescent.

11. Discuss the importance of the element carbon in the history of artificial light.

12. Look up information about some of the materials that Edison experimented with for use as filaments in the interval between the development of his paper-strip carbon and tungsten.

13. Describe the differences between photoflood and photoflash bulbs.

14. Describe the construction and operation of two types of electroluminescent sources.

15. Discuss the disadvantages for use in home lighting of electroluminescent sources, other than fluorescent lamps.

16. Get the opinion of one or more industrial users of fluorescent lighting as to its advantages and disadvantages in comparison with incandescent lighting.

17. Find as many applications as you can of fluorescent light other than those described in this chapter.

18. Distinguish between chemiluminescence and autoluminescence.

Multiple Choice

1. The intensity of illumination at a point three feet from a given source of light is approximately — (a) 1/9, (b) 1/6, (c) 1/3, (d) 3 times, (e) 6 times, (f) 9 times — that at a point one foot from the same source.

2. The results of an experiment performed with a photometer would be most likely to be expressed in which of the following units? (a) Foot-candles. (b) Candle-power. (c) Watts. (d) Per cent. (e) Volts.

3. For prolonged reading of ordinary type, the table given in this chapter recommends a minimum intensity of illumination of — (a) from 0 to 10, (b) from 10 to 20, (c) from 20 to 50, (d) from 50 to 100, (e) over 100 — foot-candles.

4. List all of the following in which incandescent carbon is the source of light. (a) Candle. (b) Sodium vapor lamp. (c) Fluorescent lamp. (d) Edison's first incandescent electric light. (e) Early arc light.

5. Which one of the following has the highest intrinsic brilliancy in terms of candle-power per square inch? (a) Kerosene lamp. (b) Prehistoric torch. (c) Bare gas flame. (d) Carbon filament. (e) Gas mantle.

True–False

1. As soon as kerosene lamps came into common use, the term "candle-power" was abandoned as the name of an intensity unit.

2. The intensity of an incandescent bulb is affected by the conditions of current and voltage at which it is operated.

3. An electroluminescent tube of blue glass containing a mixture of ›n gas and sodium vapor gives a pure white light.

4. During the nineteenth century the amount of artificial light that could obtained for a dollar increased by more than 150-fold.

5. A piece of iron that is heated, first emits visible light when it reaches emperature of between 1,500° F. and 1,800° F.

6. All the important sources of light used by man before the inven- n of the tungsten lamp were of the incandescent type.

7. In gas-filled incandescent Mazda lamps, the gas emits light much as :s neon in a neon tube.

8. Incandescent lights, on the average, have an over-all efficiency of less ın 10 per cent.

9. Increasing the temperature at which an incandescent bulb is oper- d tends to increase its efficiency.

10. In fluorescent lighting, most of the actual radiation emitted by the rcury in the tube is invisible.

Chapter 34

REFRIGERATION

SINCE energy cannot be destroyed, the cooling of a substance can be accomplished only by causing the heat energy in it to flow to some other material. In other words, the butter and milk in a refrigerator are lowered in temperature because some other substance absorbs part of the heat energy which they contained. The general methods by means of which absorption is made to occur will be our concern in the first part of this chapter, after which we shall discuss several specific types of refrigerators.

ABSORPTION OF HEAT AND CHANGE OF STATE

In connection with this part of the discussion the reader should refresh his memory concerning the phenomenon of change of state as described in Chapter 13 and illustrated in Fig. 73.

Latent Heat of Fusion

It will be recalled that while a substance is changing from a solid to a liquid it absorbs a considerable quantity of heat. In terms of the kinetic theory this heat is pictured as " pulling " the molecules of the material farther apart, thus changing its characteristics from those which we associate with solids to the group which are typical of liquids. That the heat energy does not increase the velocity of the molecules at this point is evidenced by the fact that there is no increase in temperature while the change of state is in progress. It is the fact that the amount of heat absorbed is comparatively large which gives the phenomenon its application in the field of refrigeration.

Let us consider the case of water as it changes from a solid to a liquid. One calorie of heat is required to increase the temperature of one gram of water one degree Centigrade. However, merely to melt one gram of ice requires approximately 80 calories. In other words,

the *latent heat of fusion* of ice is 80 calories per gram. Or, if we talk about ice in terms of pieces of a more familiar size, a 25-pound cake of ice absorbs almost a million calories of heat when it melts, which is almost as much heat as it takes to raise the temperature of 25 pounds of water all the way from freezing to boiling. So in the melting of ice we have available a way of getting considerable quantities of heat absorbed.

Latent Heat of Vaporization

A similar absorption of heat occurs at the other change-of-state point, that is, between liquid and vapor. Here the heat absorbed per unit weight of the vaporizing substance, called the *latent heat of vaporization*, is much larger than is the heat of fusion. In the case of water it is about 540 calories per gram or between six and seven times that required to melt a gram of ice. Thus, to vaporize 25 pounds of water requires a little over 6,000,000 calories. The cooling effect of such vaporization is a commonly experienced phenomenon. Everybody has noticed the fact that a wind feels much colder against one's hands or face if they are wet than if they are dry. This is simply because of the fact that as the water evaporates it absorbs its heat of vaporization from its surroundings. Body temperature is regulated in part by the cooling caused by the evaporation of perspiration. If it were not for the fact that a liquid absorbs its latent heat of vaporization when it changes to a vapor, an electric fan in a warm room would be useless. As it is, the fan speeds up the evaporation of perspiration by continually supplying unsaturated air to the surface of the body.

Latent Heats Related to Refrigeration

It is clear from the above description that change-of-state phenomena offer possibilities for getting large amounts of heat absorbed. If the effects are to be used as practical refrigerating procedures, several other requirements must be met.

For one thing, the material which changes state must do so at a temperature near that desired in the refrigerating device. Thus, one can employ the heat of fusion of ice because its melting point is a little but not too much below the temperature at which foodstuffs should be kept. Aluminum, for example, has to be heated to about 1,216° F.

before it melts. Obviously its change of state cannot be employed for refrigerating purposes although it has a heat of fusion about the same as that of water. For the same reason, the heat absorbed when water changes to steam cannot be used although it is more than six times as great as the heat of fusion. The evaporation of water at ordinary temperatures is employed sometimes as a cooling technique, as when jars of drinking water are wrapped in wet cloths.

A second requirement is that the substance used as a refrigerant must not affect undesirably either the materials being refrigerated, or the refrigerating device itself. Regardless of how high a heat of fusion it might have, no substance could be substituted for ice if it made foodstuffs unfit to eat or caused rapid corrosion of the ice-box itself.

Then finally, if the cooling process is to be a continuous one, some provision must be made to reverse the direction of the change of state. This means that if the refrigerating device involves the heat absorbed when a substance changes from a solid to a liquid, there must be some place where the liquid changes back into a solid again if refrigeration is to continue.

Types of Refrigerating Devices

Having seen something of the general relationship of the two kinds of latent heat to refrigeration, we shall consider next several specific types of refrigerating devices.

The Ice Refrigerator

The relative locations of ice and food in an ice refrigerator or ice-box are shown in Fig. 221. It should be noted that the ice is placed in the top compartment, above the materials that are to be kept cool. Every gram of ice that melts absorbs 80 calories of heat from the air immediately surrounding it. This cooled air, being heavier per unit volume than warm air, sinks toward the bottom of the refrigerator and warmer air flows around the ice to take its place. As the cold air moves down it absorbs heat from the butter, eggs, or meat over which it passes, lowering their temperature. Other cold air then moves down forcing the first back up to the ice compartment again. The complete circulation picture is shown by the arrows in the diagram. If the cabinet is well insulated almost all of the heat absorbed by the melting

ice comes from the food. In a poorly-insulated box much of the heat is conducted in from the outside and the ice melts without providing an adequate return to the owner in terms of the refrigeration of food. For proper food preservation the temperature inside a refrigerator should be between 40° and 50° F.

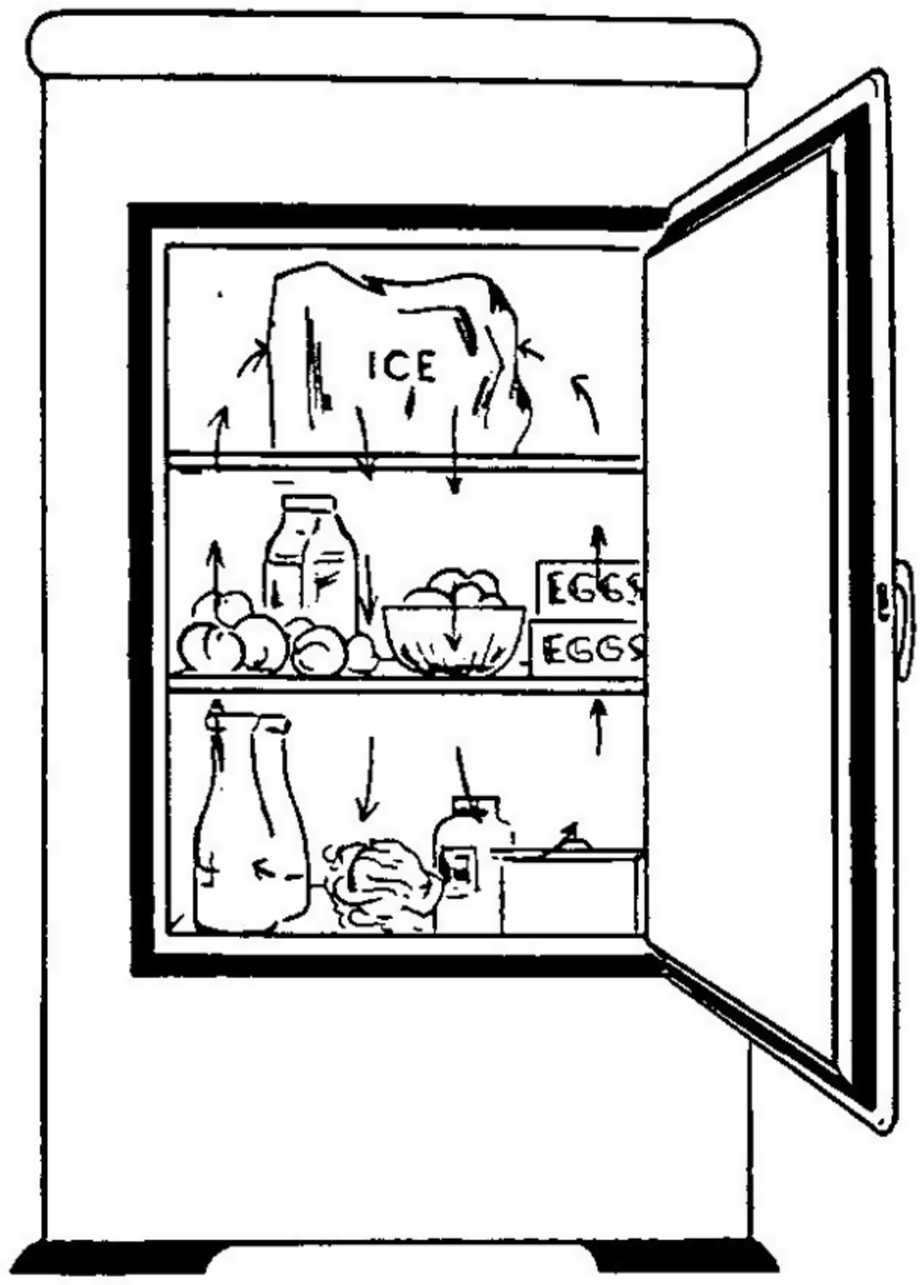


FIG. 221. Diagram of an ice refrigerator showing direction of air currents.

It should be emphasized that the device functions as an efficient refrigerator only as the ice melts and in so doing absorbs its latent heat of fusion. It is not simply an insulated container which makes it possible to place foodstuffs alongside something that is cold. If that were true no refrigerator would be worth bothering with. In the light of these facts we may ask about the practice, sometimes indulged in by housewives, of wrapping the ice in newspapers to keep it from melting so rapidly in the refrigerator. It should be clear from the above that such a procedure preserves the ice at the expense of the food. If the principal objective is to preserve the ice it would be better to let the ice company keep it, since such a firm has better ice-saving facilities than those in a kitchen refrigerator.

One comment should be made in connection with the third requirement given in the preceding section. It is obvious that only half of the complete refrigerating cycle is completed in an ice refrigerator, namely, the change from a solid state to a liquid one. The other half occurs in the plant of the ice manufacturer when water is frozen into ice. Other types of household refrigerators, as we shall see, perform the complete process.

The Compressor-Type Refrigerator

Refrigerating devices of this kind make use of the heat absorbed by a substance when it changes from a liquid to a vapor, that is, of its

latent heat of vaporization. As mentioned earlier in this chapter, a material that can be employed in this fashion must vaporize at a temperature not too far removed from the one that is to be maintained in the refrigerator.

Commercial Ice-Making Machines. — The device used in the commercial manufacture of ice commonly employs ammonia as the refrigerant. Before tracing the operating cycle of such a machine let us remind the reader of the effect that pressure has in determining whether a substance will be a liquid or a vapor at any given temperature. It will be recalled that although water ordinarily vaporizes at 212° F., it may be raised to a higher temperature as a liquid by putting it under pressure. In a pressure cooker, for example, temperatures much higher than the ordinary boiling point of water are reached without the liquid all changing to steam. But if the pressure in such circumstances is suddenly released, the water vaporizes very rapidly. The same change occurs if the super-heated water is allowed to escape through a valve into a region of lower pressure. Similarly, if liquid ammonia flows through a valve from a place of high pressure to one of low, it will change rapidly to a vapor and in so doing absorb its heat of vaporization.

Now let us trace the steps in the operation of a machine used in the manufacture of ice, as shown in Fig. 222. We will start with " Escape Valve " shown at the bottom of the figure. On the right side of this valve the system contains liquid ammonia under a pressure of from 125 to 175 pounds per square inch. As the ammonia flows through this valve in the direction indicated by the arrows, it enters a region where the pressure is around 20 pounds per square inch. This considerable decrease in pressure results in rapid vaporization of the ammonia as it flows up through the coils in the brine tank. The ammonia's heat of vaporization is absorbed from the brine, lowering its temperature under typical operating conditions to a figure as low as 5° F. This temperature is low enough to freeze the fresh water in the smaller tank shown suspended in the brine container. Thus, it is in the brine tank that the ice is actually made. But if the ammonia is to be used over again there must be some provision for converting it back into a liquid. Therefore, as the ammonia vapor leaves the left-hand coil it flows into the cylinder of a compressor through the valve marked

V_1. Here it is compressed and enters the right-hand cooling coils through V_2. In these coils, which are immersed in flowing cold water, it condenses back into a liquid. Then it is ready to escape once again through the lower valve and repeat its refrigerating performance.

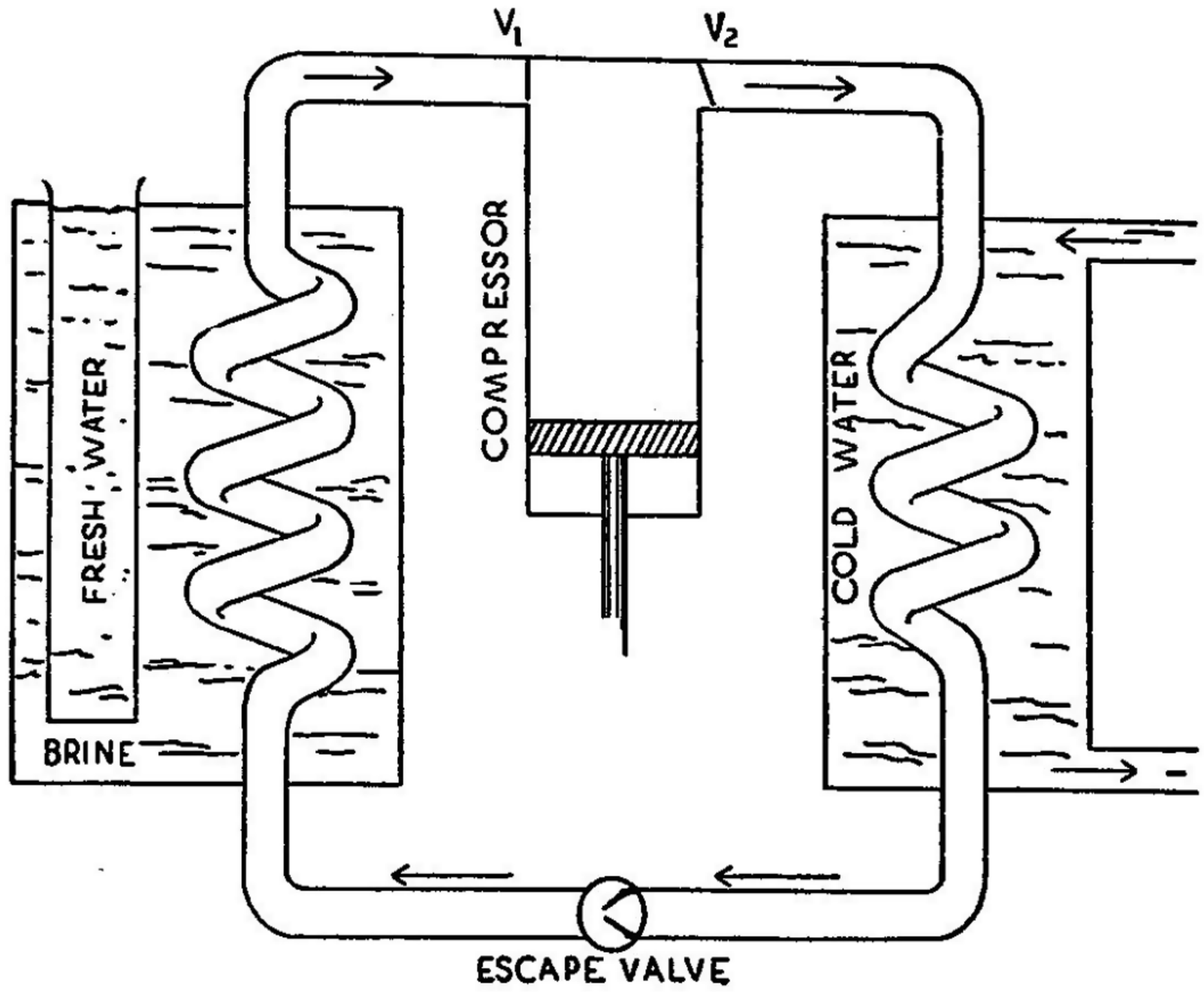


FIG. 222. Cross-sectional diagram illustrating the cycle of operation of a commercial ice-making machine.

It should be noted that the net result of the process is to remove heat energy from the brine in the left-hand tank, which is at a low temperature, and transfer it to the water in the right-hand tank which is at a higher temperature. In other words, the mechanical energy of the compressor is used to make heat flow from a lower to a higher temperature, a direction in which it does not naturally travel. This procedure is essentially the reverse of a steam engine or gas engine since in the latter case heat energy flows from a higher to a lower temperature and in so doing is made to produce mechanical energy.

Household Electrical Refrigerators. — The ordinary household electric refrigerator operates on exactly the same sort of cycle as that just described. That is to say, a liquid under pressure expands through a valve, vaporizes, absorbs its latent heat of vaporization, is compressed, and condenses back into a liquid under pressure. The most commonly used refrigerant in these machines is sulfur dioxide which is stable, non-poisonous, and changes state within the desired temperature and

pressure ranges. Others that are employed include methyl chloride, ethyl chloride, and various special compounds developed in the laboratories of refrigerator manufacturers.

In the household refrigerator the cooling coils are located around the ice-cube trays. The compressor is operated by an electric motor.

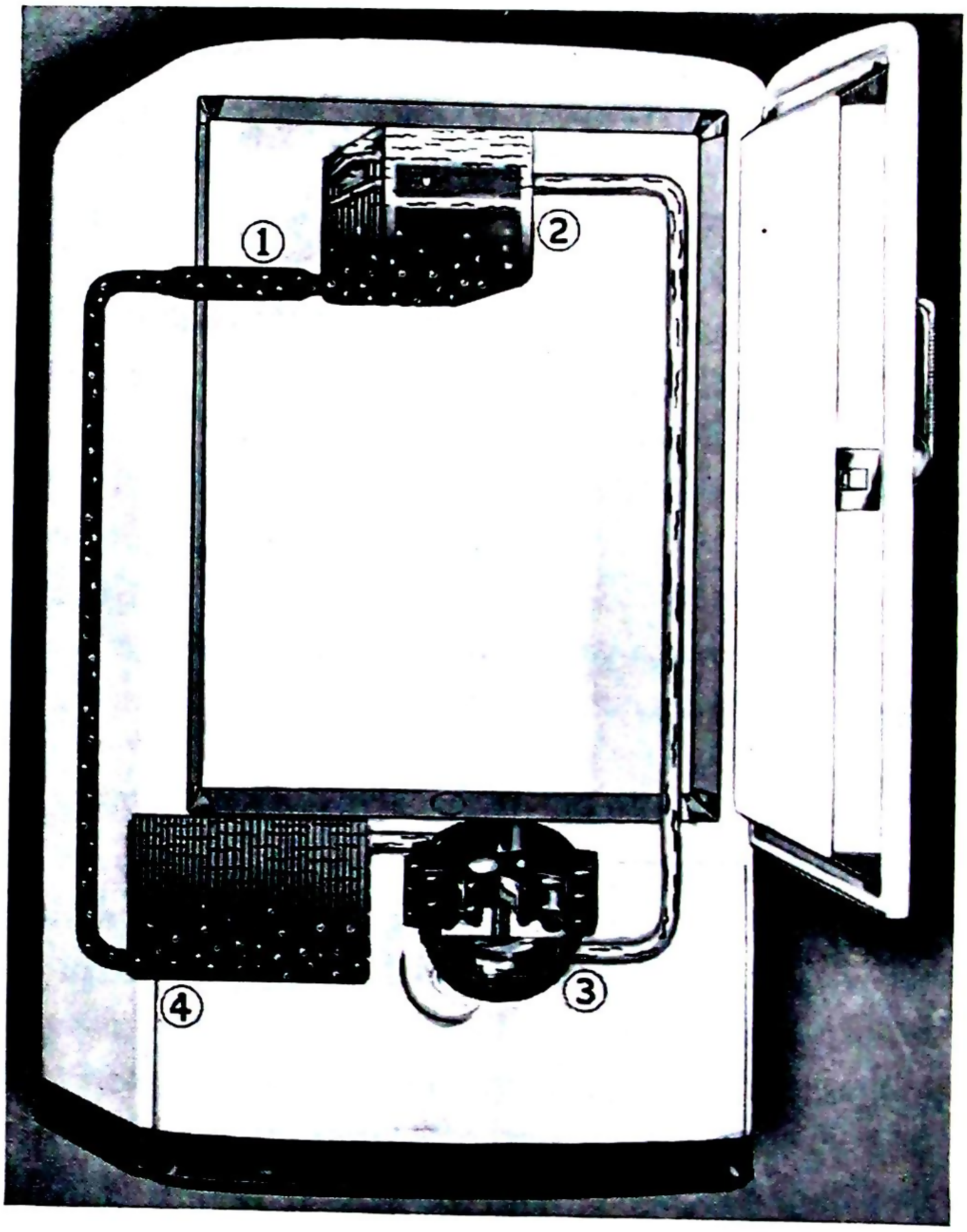


FIG. 223. Cut-away picture of an electric refrigerator. (1) Place where liquid refrigerant is released into region of lower pressure; (2) Freezing compartment; (3) Compressing pump; (4) Condenser. Liquid refrigerant is indicated by bubbles, vapor by wavy lines. (Courtesy Frigidaire.)

The coils in which the refrigerant vapor condenses back into a liquid are sometimes cooled by an electric fan. A simplified cut-away diagram of an electric refrigerator is shown in Fig. 223.

A point that should be noted in this connection is the fact that any refrigerator is also a heating device. We have already shown that

in the ice-making machine heat energy is transferred from the brine tank to the water which surrounds the condensing coils. Thus the former is cooled and the latter is heated. In the household refrigerator the heat energy that is absorbed from the foodstuff is transferred by the refrigerant around to the condensing coils from which it flows into the kitchen. In other words, the refrigeration process lowers the temperature of the materials within its walls and at the same time slightly raises the temperature of the air in the kitchen. This leads us to mention another practice in which housewives and maids occasionally indulge in summer. We refer to that of opening the refrigerator door and then ironing or doing other work in front of the device in the belief that one is keeping cool. It will be seen from the above discussion that what actually happens is that heat energy in the part of the kitchen in front of the refrigerator flows in through the open door, is then absorbed by the vaporizing sulfur dioxide or other refrigerant, next is transferred to the condensing coils in the bottom of the refrigerator, and finally goes back into the kitchen again. In other words, heat energy simply takes a round trip from the room through the refrigerator and back into the room again accompanied by a rising electric bill.

The fact that the refrigerating process results in heat flowing from a lower temperature to a higher one suggests certain possibilities in connection with heating buildings in winter and cooling them in summer. It will be recalled from the discussion in Chapter 14 of the nature of heat, that any substance whose temperature is above absolute zero contains heat energy. Thus the atmosphere, even on the coldest winter day, has a great deal of heat energy in it. Consequently, if a machine like one of the refrigerating devices just discussed were so arranged in winter that the refrigerating coils were outside and the condensing coils inside a building, heat energy could be brought into the building from outdoors. Then in summer, by reversing the process, heat energy could be transferred from indoors to outdoors, thus using the device to cool the building.

The Gas Refrigerator

Another type of refrigerating device, known as the " gas refrigerator," is in common household use today. It receives its name from

the fact that a gas flame is kept burning under one part of the mechanism. Its operation is based on a discovery made by two Swedish boys, Carl Münters and Baltzar von Platen, while they were undergraduates in the Royal Institute of Technology at Stockholm.

Like the electric refrigerator, the gas type makes use of the heat absorbed by a liquid when it changes to a vapor, ammonia being the refrigerant employed in this case. Its chief variation from the compressor-type machine is in the manner in which the second part of the refrigerating cycle is accomplished, that is, the change from a vapor back into a liquid again. This will be evident as we trace through the cycle of operation as pictured in Fig. 224, which illustrates the various circuits in this type of refrigerator. First, it should be noted that three substances are enclosed in the system, namely, ammonia, water, and hydrogen. Each one follows its own circuit through the device. Let us begin the story of the operation of this device at the *generator*, shown at the bottom of the figure. There flows into the generator a solution of ammonia in water. The gas flame, which burns continuously under this part of the outfit, drives the ammonia out of solution in vapor form and causes both it and the water to rise through the vertical pipe to the *separator*. This rise occurs in much the same manner that liquid flows up through the central stem in a coffee percolator. From the separator the water flows by gravity to the absorber while the ammonia vapor moves on up to the *condenser*. Here, as the name suggests, the ammonia condenses into a liquid. In earlier models of this refrigerator it was necessary to have a small stream of water flowing over the condenser at all times in order to cool the ammonia vapor enough so that it would liquefy. In recent machines, sufficient lowering of temperature is obtained by simply having around the condenser metal fins which radiate the heat. The liquid ammonia next flows to the left and into the *evaporator* where it encounters an atmosphere of hydrogen into which it evaporates, thus absorbing its heat of vaporization and doing the work of refrigeration. It will be recalled that in the compressor-type machines the coils in which vaporization occurs contain simply air. The use of hydrogen, however, has certain definite advantages because of its lightness, its insolubility in water, and the fact that it does not corrode the metal tubes. The mixture of ammonia gas and hydrogen next moves down from the evaporator and

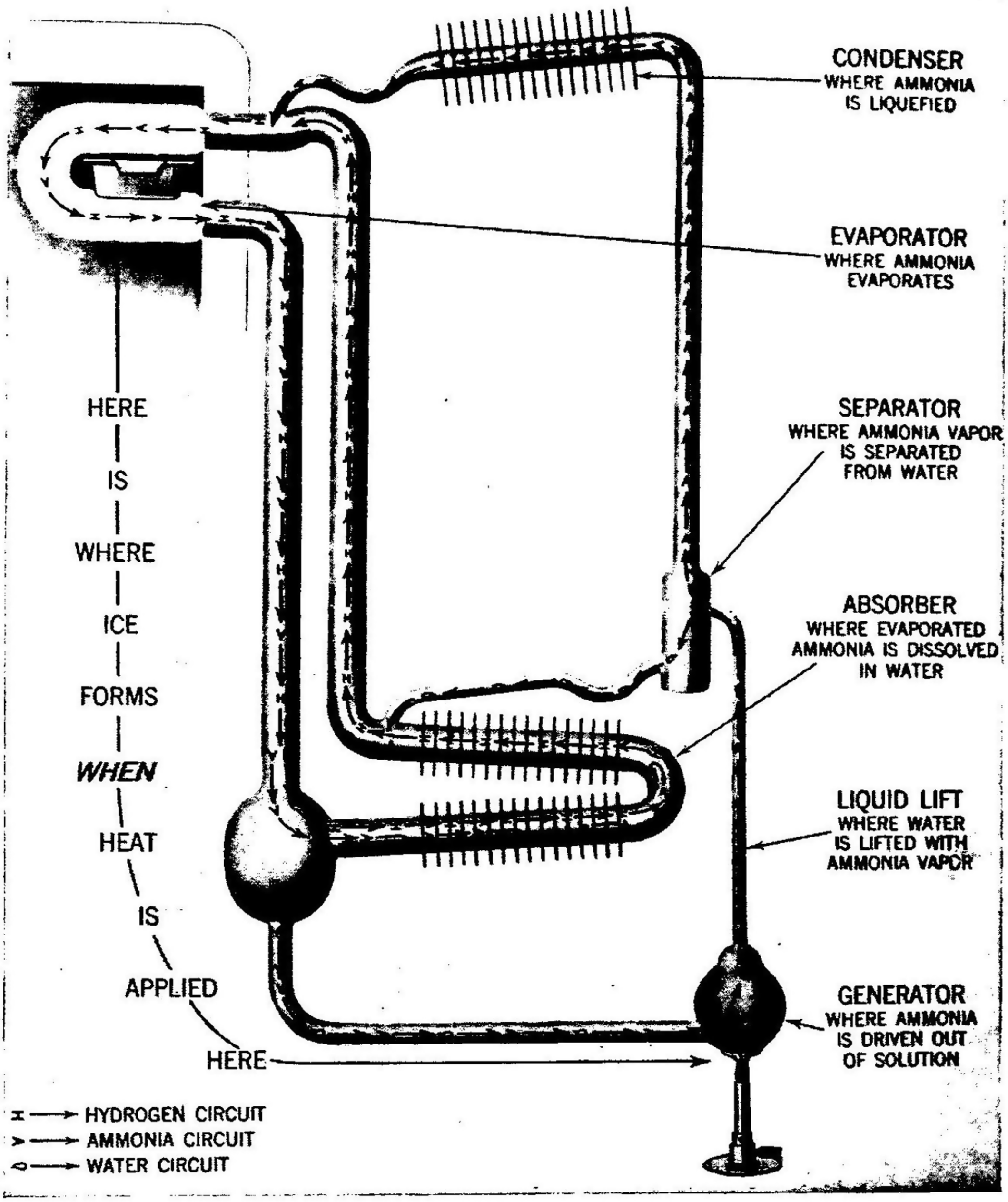


FIG. 224. Diagram showing the various circuits in the system of a gas refrigerator. (Courtesy Servel-Electrolux.)

into the *absorber*. Here it encounters the water which, as was mentioned above, flows down from the separator on the other side. The water serves the purpose of separating the ammonia and hydrogen since the former will dissolve in water but the latter will not. The solution of ammonia in water then flows over to the generator while

the hydrogen rises from the absorber up through the vertical pipe to join more ammonia entering the evaporator. Convection currents in a gas contained in the separate tube which surrounds the absorber serve to cool this part of the system.

The amount of refrigeration produced in this type of refrigerator is governed by the amount of heating at the generator. Turning up the gas flame results in more rapid evaporation taking place in the evaporator and consequently greater cooling is accomplished. Since the process is a continuous one rather than one which stops and starts at intervals, a high degree of constancy of temperature is obtainable in the refrigerator.

It will be noted that although the device is commonly called a gas refrigerator any means of heating the generator might be expected to prove satisfactory. Both kerosene burners and electric devices are possible sources of heat.

Questions

Discussion

1. Define the latent heats of fusion and of vaporization and distinguish between them.

2. Describe the experiment the results of which are shown in Fig. 73 in Chapter 13.

3. In terms of the kinetic theory of matter and heat, what is the physical significance of each of the following experimental facts? (a) Under certain conditions materials increase in temperature as they absorb heat. (b) Under certain conditions, materials change state as they absorb heat.

4. Compare the average velocity of the molecules of ice at 32° F. with that of water molecules at 32° F. Explain.

5. Which would be the better container for drinking water on a very hot dry day, a semi-porous earthen jar or a glazed porcelain one? Explain.

6. Explain the cooling effect of an electric fan operating in a warm closed room.

7. Explain why neither the heat of fusion of iron nor the heat of vaporization of water is available for use in connection with commercial refrigeration.

8. Name and discuss the requirements for a satisfactory refrigerant as they were listed in this chapter.

9. Why would a refrigerator with the ice compartment at the bottom be less efficient than one with it at the top?

10. If the cost were the same would a 25-pound block of glass at 32° F.

be more, or less desirable to have in an ice-box than a 25-pound cake of ice at 32° F.? Explain.

11. Draw a diagram of the system in a commercial ice-making machine and explain its operation.

12. What becomes of the electrical energy that enters an electrical household refrigerator? Trace its transformations step by step.

13. Discuss the relation of pressure to the vaporization of a liquid.

14. Explain just how an electrical refrigerator helps to heat the room in which it is located.

15. Other things being equal, should leaving the refrigerator door open for several hours cool the room off, warm it up, or have no effect on it? Explain.

16. Contrast the refrigerating processes in the electrical and gas refrigerators.

17. Turning up the gas under the generator in a gas refrigerator is analogous to what procedure in the electrical refrigerator?

Multiple Choice

1. To change 1 gram of ice at 0° C. into steam at 100° C. requires the absorption of about — (a) 80, (b) 100, (c) 540, (d) 640, (e) 720 — calories of heat.

2. Which one of the following facts has the most direct bearing on the reason for putting the ice at the top of the ice refrigerator rather than at the bottom? (a) Ice contracts as it melts. (b) Over six times as much heat is required to vaporize a pound of water as it takes to melt a pound of ice. (c) A cubic foot of cold air weighs more than a cubic foot of warm air. (d) A cubic foot of ice weighs less than a cubic foot of water. (e) The melting point of ice lowers as the pressure on it is increased.

3. List all of the following in which the latent heat of vaporization of some substance is absorbed during the refrigeration process. (a) The gas refrigerator. (b) The household electric refrigerator. (c) Commercial ice-making. (d) The ice refrigerator.

4. In the commercial ice-making machine the pressure in the condensing coils is between — (a) 0 and 5, (b) 6 and 10, (c) 11 and 15, (d) 16 and 20, (e) 21 and 30 — times that in the cooling coils.

5. List all of the following substances that follow a circuit through the system of a gas refrigerator. (a) Oxygen. (b) Ammonia. (c) Sulfur dioxide. (d) Hydrogen. (e) Water.

True–False

1. As a substance absorbs heat it either increases in temperature or changes state.

2. The walls of a refrigerator should be made of substances that are good

conductors of heat in order to conduct heat energy to the outside as rapidly as possible.

3. Other things being equal, decreasing the rate at which ice melts in a refrigerator decreases the amount of refrigeration obtained in a given time.

4. A liquid that vaporized at 90° C. would be a poor refrigerant even though its latent heat of vaporization might be very large.

5. In any electrical or gas refrigerator one part of the system is being heated while another is being cooled.

6. In view of its use in making ice one may infer that brine has a higher freezing point than pure water.

7. The greater the pressure on a liquid, the lower is its boiling point.

8. Any considerable amount of refrigeration is obtained from an ice-box, only if the ice melts.

9. In the general refrigeration process heat energy is made to flow from a point of lower temperature to one of higher.

10. All three materials in the system of a gas refrigerator follow exactly the same circuit through the device.

READING LIST

AUTHOR	TITLE	READ IN CONNECTION WITH THE FOLLOWING CHAPTERS:
Albert, A. L.	*Electrical Communication.* New York: John Wiley, 1934.	32.
Allen, J. S., et al.	*Atoms, Rocks, and Galaxies.* New York: Harper, 1938.	General.
Amer. Council on Educ.	*The Story of Numbers.* Pamphlet, 1932.	11.
	The Story of Our Calendar. Pamphlet.	12.
	Telling Time Through the Centuries. Pamphlet.	12.
	Story of Weights and Measures. Pamphlet.	12.
Anderson, E. P.	*Electronic Devices and Their Application.* New York: Theo. Audel & Co., 1938.	27.
Anderson, J. W.	*Diesel Engines.* New York: McGraw-Hill, 1935.	22.
Baker, R. H.	*Astronomy.* New York: Van Nostrand, 1938.	2, 3, 4, 5.
Barker, P. W.	*Rubber.* U. S. Dept. of Commerce.	18, 19.
Bartky, W.	*Highlights of Astronomy.* Chicago: Univ. of Chicago, 1935.	2, 3, 4, 5, 6.
Bawden, A. T.	*Man's Physical Universe.* New York: Macmillan, 1943.	General.

AUTHOR	TITLE	READ IN CONNECTION WITH THE FOLLOWING CHAPTERS:
Beery, P.	*Stuff.* New York: Appleton-Century, 1936.	15, 16, 17.
Black, N. H.	*Introductory Course in College Physics.* New York: Macmillan, 1941.	22, 33.
Blackwood, O.	*Introduction to College Physics.* New York: John Wiley, 1938.	25.
Bradley, J. H.	*Autobiography of the Earth.* New York: Coward-McCann, 1935.	7, 8, 9, 10.
Bragdon, L.	*Words and Wings.* New York: Farrar & Rinehart, 1938.	32.
Branson and Tarr	*An Introduction to Geology.* New York: McGraw-Hill, 1935.	7, 8, 9, 10.
Brinkley, S. R.	*Introduction to General Chemistry.* New York: Macmillan, 1938.	17, 18.
Bush, Ptacek, and Kovats	*Senior Science.* New York: American Book, 1937.	13, 14, 15, 16, 34.
Cable, Getchell, and Kadesch	*The Physical Sciences.* New York: Prentice-Hall, 1940.	General.
Cajori, F.	*A History of Mathematics.* New York: Macmillan, 1919.	11.
Caldwell, F. C.	*Modern Lighting.* New York: Macmillan, 1930.	33.
Cottman, E.	"Cold Light." *Scientific American*, June 1939.	33.

AUTHOR	TITLE	READ IN CONNECTION WITH THE FOLLOWING CHAPTERS:
Cramer, H. I.	"Industrial Progress in Synthetic Rubber-like Polymers." *Ind. & Eng. Chemistry*, Feb. 1942.	19.
Crew, H.	*The Rise of Modern Physics.* New York: Williams & Wilkins, 1935.	13, 14.
Croneis and Krumbein	*Down to Earth.* Chicago: Univ. of Chicago, 1936.	7, 8, 9, 10.
Dampier, Sir Wm.	*A History of Science.* New York: Macmillan, 1930.	General.
DeMent and Dake	*Uranium and Atomic Power.* Cleveland: Chemical Publishing Co., 1945.	26.
Desha, L. J.	*Organic Chemistry.* New York: McGraw-Hill, 1936.	16.
Dickinson, H. W.	*A Short History of the Steam Engine.* New York: Macmillan, 1941.	22.
Dietz, David	*Atomic Energy in the Coming Era.* New York: Dodd, Mead, 1945.	26.
Draffin, J. O.	*The Story of Man's Quest for Water.* Champaign, Ill.: Garrard Press, 1939.	15.
Ely, S. B.	"The Diesel Engine and Its Possibilities." *Scientific Monthly*, April 1937.	22.
Encyclopaedia Britannica, Ninth Edition.		32.
Eyring, C. F.	*A Survey Course in Physics.* New York: Prentice-Hall, 1936.	21, 25, 28, 29, 30, 31, 34.

AUTHOR	TITLE	READ IN CONNECTION WITH THE FOLLOWING CHAPTERS:
Fisher, H. L.	*Rubber — Its Uses.* Cleveland: Chemical Rubber Co., 1941.	18, 19.
Fisk, D.	*Exploring the Upper Atmosphere.* New York: Oxford Univ. Press, 1934.	7.
Foley, A. L.	*College Physics.* Philadelphia: Blakiston, 1941.	13, 14, 22, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34.
Ford Motor Co.	*History of Measurement of Light.* Pamphlet.	12.
Fortune	"Plastics." March 1936.	19.
Foster, W.	*The Romance of Chemistry.* New York: Century, 1927.	15, 16, 17, 18, 19.
——	*Inorganic Chemistry for Colleges.* New York: Van Nostrand, 1936.	17.
Furnas, C. C.	*The Next Hundred Years.* New York: Reynal & Hitchcock, 1936.	7, 21, 22, 33.
General Motors Corp.	*When the Wheels Revolve.* Facts about Automobiles. Pamphlet.	22.
	Diesel, The Modern Power. Pamphlet.	22.
	Electricity and Wheels. Pamphlet.	23, 25.
	Optics and Wheels. Pamphlet.	30.
George and Gilman	*Modern Mercuries.* New York: Robt. McBride, 1932.	32.
Gillmer and Nietsch	*Clouds, Weather and Flight.* New York: Van Nostrand, 1944.	7.

AUTHOR	TITLE	READ IN CONNECTION WITH THE FOLLOWING CHAPTERS:
Ginzburg, B.	*The Adventure of Science.* New York: Simon & Schuster, 1930.	15, 24, 25.
Grieveson, C. J. W.	*Symposium on Illumination.* Pittsburgh, Pa.: Instruments Publ., 1935.	33.
Harrington, J.	"Seeing with Electricity." *Scientific American,* July 1940.	33.
Hausmann and Slack	*Physics.* New York: Van Nostrand, 1939.	22, 26, 27, 28, 29, 30, 31.
Hawley and Leifson	*Atomic Energy in War and Peace.* New York: Reinhold Company, 1945.	26.
Heil, L. M.	*The Physical World.* Philadelphia: Blakiston, 1936.	General.
Hildebrand, J. R.	"Our Most Versatile Vegetable Product." *National Geographic,* February 1940.	18.
Hoag, J. B.	*Basic Radio.* New York: Van Nostrand, 1942.	27.
Hobbs, W. H.	*Earth's Features and Their Meaning.* New York: Macmillan, 1935.	7, 8.
Hogben, L.	*Science for the Citizen.* New York: Knopf, 1938.	General.
——	*Mathematics for the Millions.* New York: Norton, 1937.	11.
Holmes, H. N.	*Introduction to College Chemistry.* New York: Macmillan, 1939.	17, 18.

AUTHOR	TITLE	READ IN CONNEC WITH THE FOLI ING CHAPTER
Jaffe, B.	*Outposts of Science.* New York: Simon & Schuster, 1935.	7.
Jean, Harrah, Herman, and Powers	*Man and the Nature of His Physical Universe.* Boston: Ginn, 1934.	8, 20, 32.
Johnson, J.	"Applications of Science to Metallurgical Industry." *Scientific Monthly*, June 1939.	17.
Kimball, A. L.	*College Physics.* New York: Henry Holt, 1923.	20.
Kunnerth, W.	*Textbook on Illumination.* New York: John Wiley, 1939.	33.
Landis, W. B.	"Rubber for America." *Scientific American*, November 1940.	18.
Lee, R. E.	*Backgrounds and Foundations of Modern Science.* New York: Macmillan, 1938.	21.
Lemon, H.	*From Galileo to Cosmic Rays.* Chicago: Univ. of Chicago, 1934.	14, 15, 20, 21 26, 28, 29, 30
Lind, W. E.	*Internal Combustion Engines.* Boston: Ginn, 1920.	22.
Loeb and Adams	*Development of Physical Thought.* New York: John Wiley, 1933.	1, 2.
Logsden, M.	*A Mathematician Explains.* Chicago: Univ. of Chicago, 1936.	11.
Luckiesh, M.	*Artificial Light.* New York: Century, 1920.	33.
Luckiesh and Moss	*The Science of Seeing.* New York: Van Nostrand, 1937.	33.

AUTHOR	TITLE	READ IN CONNECTION WITH THE FOLLOWING CHAPTERS:
Manchester, H.	"Daylight in Tubes." *Atlantic Monthly*, July 1941.	33.
McCorkle, P.	*Survey of Physical Sciences.* Philadelphia: Blakiston, 1938.	General.
Menzel, D.	*Stars and Planets.* New York: Univ. Soc., 1931.	2, 3, 4, 5, 6.
Mills, J. S.	*Electronics Today and Tomorrow.* New York: Van Nostrand, 1944.	27.
Mott-Smith, M.	*The Story of Energy.* New York: Appleton-Century, 1934.	22.
Moulton, F. R.	*The World and Man.* Chicago: Univ. of Chicago, 1937.	7, 8, 9, 10.
Moulton and Schifferes	*Autobiography of Science.* Garden City, N. Y.: Doubleday, Doran, 1945.	General.
Newman et al.	*The Nature of the World and Man.* Garden City, N. Y.: Garden City Press, 1933.	7.
Olcott, W. T.	*Star Lore.* New York: Putnam's, 1911.	4.
Osborn, F. A.	*Physics of the Home.* New York: McGraw-Hill, 1925.	15, 30, 31, 32, 33. 34.
Peck, A. P.	"What Light for Industry?" *Scientific American*, May 1941.	33.
Pendray, G. Edward	*The Coming Age of Rocket Power.* New York: Harper & Brothers, 1945.	22.

AUTHOR	TITLE	READ IN CONNECTION WITH THE FOLLOWING CHAPTERS
Polya, G.	*How to Solve It.* Princeton: Princeton Univ. Press, 1945.	11.
Reeds, C. A.	*The Earth.* New York: Univ. Soc., 1931.	7.
Richards, H.	*The Universe Surveyed.* New York: Van Nostrand, 1937.	General.
Richardson, E. G.	*Physical Science in Modern Life.* New York: Van Nostrand, 1939.	13, 14.
Robertson, John K.	*Atomic Artillery and Atom Bomb.* New York: Van Nostrand, 1945.	26.
Sabine, P. E.	*Acoustics and Architecture.* New York: McGraw-Hill, 1932.	29.
Schonland, B. F. J.	*Scientific Monthly*, April 1936.	24.
Servel, Inc.	*The Miracle of Gas from Heat.* Pamphlet.	34.
Shaw, N.	*The Drama of the Weather.* New York: Macmillan, 1934.	7.
Simonds, Herbert R.	*The New Plastics.* New York: Van Nostrand, 1943.	19.
Simonds and Ellis	*Handbook of Plastics.* New York: Van Nostrand, 1942.	19.
Smith, G. Geoffrey	*Gas Turbines and Jet Propulsion for Aircraft.* New York: Aerosphere, Inc., 1944.	22.

AUTHOR	TITLE	READ IN CONNECTION WITH THE FOLLOWING CHAPTERS:
Smith, T. A.	"The Electron Microscope." *Scientific Monthly*, April 1941.	31.
Smith and Ginzburg	*Numbers and Numerals.* Pamphlet, 1937.	11.
Smyth, H. D.	*Atomic Energy for Military Purposes.* Princeton: Princeton Univ. Press, 1945. Washington: Government Printing Office, 1945.	26.
Snider, L. C.	*Earth History.* New York: Century, 1932.	7, 8, 9, 10.
Stetson, H. T.	"The Influence of the Sun on Human Affairs." *Scientific Monthly*, July 1936.	2.
Stewart, John Q.	*Coasts, Waves and Weather.* Boston: Ginn, 1945.	7.
Stewart, O. M.	*Physics.* New York: Ginn, 1939.	20, 28.
Stommel, Henry	*Science of the Seven Seas.* New York: Cornell Maritime Press, 1945.	General.
Swann, W. F. G.	*Architecture of the Universe.* New York: Macmillan, 1934.	3.
Sweinhart, J.	*He Measures in Millionths.* Ford Pamphlet.	12.
Taylor, L. W.	*Physics, the Pioneer Science.* New York: Houghton Mifflin, 1940.	General.
Thomson, J. A.	*Outline of Science.* New York: Putnam, 1937.	3, 4, 5.
Timm, J. A.	*An Introduction to Chemistry.* New York: McGraw-Hill, 1932.	1, 13, 14, 15, 16, 17, 18, 19.

AUTHOR	TITLE	READ IN CONNECTION WITH THE FOLLOWING CHAPTERS
Tucker, Durward J.	*Introduction to Practical Radio.* New York: Macmillan, 1945.	27.
Wallace and Tiernan Co.	*Technical Reprint* # 467. Reprinted from Annual Yearbook, American Public Health Assn. Vol. 20, #2.	15.
Watkeys et al.	*An Orientation in Science.* New York: McGraw-Hill, 1938.	4, 7, 8, 9, 10, 17, 20.
Weeks, Mary E.	"Discovery of the Elements." Lancaster: *Journal of Chemical Education,* 1945.	13, 14.
Wells, H. G.	*Outline of History.* New York: Macmillan, 1921.	1.
Wells, Huxley, Wells	*The Science of Life.* Garden City, N. Y.: Garden City Press, 1939.	9, 10.
Westaway, F. W.	*The Endless Quest.* New York: Hillman-Curl, 1936.	General.
White, A. D.	*History of Warfare Between Science and Theology.* New York: Appleton-Century, 1888.	9.
White, H. E.	*Classical and Modern Physics.* New York: Van Nostrand, 1940.	13, 14, 15, 23, 25, 26, 27, 28, 30, 31, 32, 33,
Wolf, A.	*A History of Science, Technology and Philosophy in the 16th and 17th Centuries.* New York: Macmillan, 1936.	22.
Wolf, Howard and Ralph	*Rubber.* New York: Covici, Friede, Inc., 1936.	18.
Wood, A. B.	*A Textbook of Sound.* New York: Macmillan, 1932.	29.
Yates, Raymond F.	*Atom Smashers.* New York: Didier Press, 1945.	26.

Index

(The presence of an * after an item indicates that there is a figure related to it)

Absolute temperature scale, 254–255 *
Absolute zero, 255
Absorption of heat, 243
Absorption spectrum, 81
Acadian Mountains, 180
Acceleration, 371–373
Accommodation, 608
Acetylene gas, 292–293
Acoustics, 551–555 *
Acryloids, 358–359
Actinium, 495
Activity series, chemical, 297, 299
Adams, John Couch, 40
Addition in Roman numerals, 197
Aeolipyle, 417
Age of earth, 91, 154–157
Age of Fuels, 278
Age of Man, 183
Age of Reptiles, 182
Air (see Atmosphere also),
  density, 260
  pressure, 260–263
  properties as a liquid, 263–264
Aircraft detection, 603
Air-mass analysis, 125–126
Airplane engines, 412–413 *
Airstream engine, 417, 418–422 *
Alcohol, 290
Alexander the Great, 8
Algae, 179
Alleghany Mountains, 138
Alloys,
  table of, 315–317
  types, 313–314
Alnico, 313, 440
Alpha Centauri, 22, 23, 36, 65, 82
Alpha rays, 496
Alps, 144, 182
Alternating current, 469
Altimeter, 262–263
Aluminum,
  general discussion, 309–311
  in earth's crust, 114
Ammonia, 352–353
Ampère, André, 467
Ampere of current, 472, 473
Amphibians, 180, 181
Amplitude,
  of sound waves, 549–550
  of waves in general, 541–542 *
Anaximenes, 145
Ancient astronomy, 16–18 *
Andes Mountains, 144, 182
Andromeda nebula, 66
Aneroid barometer, 262
Anode deposition of rubber, 338
Antarctic circle, 98 *
Antares, 65
Anthracite coal, 283
Anticyclone, 124
Anti-knock gasoline, 409
Appalachia, 180
Appalachian Mountains, 144, 181
Applied Physics Laboratory, Johns Hopkins University, 422
Arabic numerals (see Hindu-Arabic numerals)
Arabs, 10, 195
Arcas, 63
Archeozoic era, 176, 179
Archimedes, 8–9
Archimedes principle, 9
Arc light, 466
Arctic circle, 98 *
Arcturus, 64–65
Argon, 116, 266
Aristarchus, 16, 17
Aristotle, 8, 63, 220, 296, 592 *
Artificial fever, 601–602

Assigned radio range, 599, 600–601
Asteroids (see Planetoids)
Astigmatism, 609 *
Astronomical measurements, 70–86
Astronomy, History of, 16–22
Athodyd, 422 *
Atmosphere (see Air also),
  composition, 116
  extent, 116
  stratification, 116–119
Atmospheric circulation,
  and earth's rotation, 120
  and weather, 119–120
Atmospheric pressure, 260–261
Atom,
  definition, 222
  nuclear model, 499–503
Atomic,
  building blocks, 499
  structure, 484–503 *
  theory, early history, 221
  weight, 223–225
Atomic bomb, 508–509
Atomic energy, 503–510
Atom smashing, 505–507
Auditorium acoustics, 551–555 *
Aurora australis, 116, 119
Aurora borealis, 116, 119
Autoluminescence, 654
Automobile,
  engine (see Gas engine)
  ignition, 406
Average solar day, 210
Axis, earth's, 93
Azoic era, 174, 176, 179

Babbit metal, 312, 315
Babylonian calendar, 211
Babylonians, 16
Bacon, Francis, 271
Barometer, 261–262 *
Basalt, 131
Batch distillation, 288
Batteries, electric (see Cells, electric)
Bauxite, 310
Beats, 557–558 *
Becquerel, Henri, 494
Becquerel rays, 495
Bedford limestone, 181
Beehive coke oven, 284
Bell, Alexander Graham, 628
Berea sandstone, 181
Bergius, Friedrich, 289
Bessemer, Sir Henry, 306
Bessemer steel, 306 *
Beta rays, 496 *
Betelgeuse, 25
Bible,
  and age of earth, 155
  and model of universe, 20
Big Dipper, 62–64, 95 *
Bikini Atoll, 511
Binding energy, 505
Binoculars, 610
Bituminous coal, 283
Black Hills, 149–150, 179, 180
Blast furnace, 302–305 *
Block and tackle, 379
Bohr, Niels, 500
Bolometer, 241
Boötes, 64
Boyle's law, 234, 253 *
Brachiopods, 180
Brahe, Tycho, 10
Brass, 296, 309, 312, 315
Bright-line spectrum, 80
British thermal unit, 243
Bromine, 270
Bronze, 295, 315
Bronze Age, 6, 295
Brown, H. R., 281
Brown, Robert, 256
Brownian movements, 256, 323 *
Burning (see Combustion)
Burning glass, 72 *
Butadiene, 344
Butyl rubber, 346
Buzz-bomb, 421
By-product coke oven, 284–285
By-products,
  of charcoal manufacture, 286
  of coke manufacture, 285

Caesar, Augustus, 213
Caesar, Julius, 104, 213

Calcium carbide, 292–293
Calendars,
  ancient, 211–212
  Gregorian, 213–214
  Julian, 212–213
  proposed new, 214
Calendering of rubber, 334
Callisto, 63
Caloric theory of heat, 246–249
Calorie, 243
Cambrian period, 180
Camera, 607 *
Canadian Shield, 179
Canals on Mars, 37
Cancer, Tropic of, 97 *
Candle-power, 636–637
Capricorn, Tropic of, 98 *
Carbon, 350, 351
Carbon dioxide, 116, 265–266
Carbon monoxide, 282, 304
Carbonation and erosion, 141
Carbonization and fossils, 162
Carburetor, 406
Carlsbad Caverns, 138
Carnegie Institution, 522
Cascadia, 180
Cassiopeia, 64
Cast iron, 305
Cathode rays, 486–488
Cathode ray tube, 486 *
Catskills, 138
Cavendish, Henry, 269, 274–275
Celanese, 361
Cells, electric, 461–465 *
Celluloid, 362–363
Cellulose acetate, 361
Cellulose nitrate, 362
Cenozoic era, 176, 182–183
Centigrade temperature scale, 241–242 *
Centrosphere, 111–113
Ceres, 41
Chain reaction, 506
Chamberlain, T. C., 48
Change of state, 235–237
Characteristic spectrum, 80
Charcoal, 286
Charged bodies, 451–453
Chemical activity series, 297, 299
Chemical equations, 229–230
Chemical formulae, 228
Chemical symbols, 223
Chemiluminescence, 653–654
Chicago World's Fair, 65
Chlorination of water, 272
Chloroprene, 344
Chromosphere, 26 *
Circumpolar whirls, 120
Cleveland Weather Bureau, 123
Climate, 119, 170, 171, 172
Cloud chamber, Wilson, 496–497
Clouds, 121
Cloudy suspension, 321
Clues to earth history, 154–166
Coagulation of latex, 330–331
Coal, 282–284
Coal gas, 285, 291
Coal oil (see Kerosene)
" Coal Sack " nebula, 60 *
Coal tar products, 285
Coarse suspension, 321
Coke, 284–286
Colloids,
  examples, 322
  properties, 322–325
  size of particles, 319–321
  types, 320–322
Color,
  nature of, 613–616
  of mixed light, 615–616
  of mixed pigments, 616
  production of, 617–619
Colors, complementary, 616
Color vision, 617
Columbia plateau, 150–151
Combustion,
  and concentration, 281–282
  and surface, 280–281
  defined, 279
  effect of temperature on, 279–280
  spontaneous, 279–280
Cometary orbits, 42
Comets, 42–43
Commercial broadcasting, 600
Communication, 622–636

Communication methods,
printing, 624–626
radio, 630–632
speech, 623
telegraph, 626–627
telephone, 627–630
writing, 623
Compass, 430–431
Complementary colors, 616
Compounds, 228–231
Compressibility of,
gases, 233, 251
liquids, 232, 251
solids, 231, 251
Compressional waves, 538–539
Compression ratio, 409
Concentration of ore, 299–300
Condensation, 545
Conduction of heat, 244–245 *
Conformity, 158
Conservation of energy, 390–391, 392
Constellations, 61
Continental glacier, 139
Continental shelves, 114
Continuous distillation, 288
Continuous duct engine, 422 *
Continuous spectrum, 80
Control grid, 517
Controlled experiment, 5
Convection, 170, 245 *
Copernican universe, 19
Copernicus, Nikolaus, 10, 18–19
Copolymerization, 345
Copper, 308–309
Copper plating, 478
Coral, 165–166
Corona, 28, 106 *
Corrosion, 297
Cosmic rays, 606
Coster, Laurens, 625
Cottrell precipitation, 324 *
Coulomb, 472, 473
Counting by early man, 189–190
Cracking of gasoline, 289
Crankshaft, 401
Craters on moon, 100
Crest of wave, 537
Cretaceous period, 182
Critical angle, 584
Crookes, Sir William, 486
Crookes tube, 486–488 *
Crust of earth, 112, 113–115
Cryolite, 310
Crystalline solid, 329
Cubit, 203
Cunningham, Leland S., 43
Cunningham comet, 43 *
Curie, Marie and Pierre, 13, 494,
Current electricity, 459–481
Cycles,
four-stroke, 410
geological, 169–173
radio, 600
Cyclone, 123

DDT, 365
Dalton, John, 221, 222, 223
Dampier, Sir William, 220
Dark-line spectrum, 80
Dark nebulae, 59–60 *
Da Vinci, Leonardo, 10 *
Davy, Sir Humphry, 248, 390, 64
Davy safety lamp, 280
Dawn Man, 183
Day and night,
cause of, 93
comparative lengths of, 98–99
Day, average solar, 210
Days of week, 209
Decimal fractions, 198
Declination, magnetic, 431 *
Definite composition, law of, 228
De-icers, 340
De Laval, Carl, 403
De-magnetizing methods, 434–4
438–439
Democritus, 145, 220
Derived units of measurement, 21
Destructive distillation, 284
Deuterium, 274
Deville, 310
Devil's Tower, 151
Devonian period, 180
Dew, 122
Dew point, 121

Diathermy machines, 602
Diatomic gases, 229
Diatonic scale, 565–567
Diderot, 187, 200
Diesel engine, 413–416
Diffraction of light, 575–577 *
Diffuse nebula, 59 *
Diffusion,
  of gases, 233, 251
  of liquids, 232, 251
  of solids, 231–232, 251
Digits, 191
Dikes, 149
Dinosaur fossils, 162
Dinosaurs, 181, 182 *
Dip, magnetic, 432 *
Dipping needle, 432 *
Direct proportion, 234
Displacement reaction, 298–299
Distillation,
  destructive, 284
  fractional, 263
Doldrums, 120, 122
Dome mountains, 149
Doppler effect,
  and star velocities, 85
  in light, 587
  in sound, 559–560 *
Double stars, 58
Douglas, Andrew, 28
Dow Chemical Co., 270
Dry cell, 464
" Dry ice," 235, 266
Ductility, 296
Dufay, Charles, 446–447
Duodecimal system, 191
Duprene (see Neoprene)
Dust explosions, 281
Dwarf stars, 57
Dynamo, 468–470
Dyne, 374

Earth,
  age of, 91–92, 154–157
  and habitability, 92–93
  and weather, 120
  area of surface, 115
  as a magnet, 429–432
Earth — *continued*
  as an astronomical body, 89–129
  circumference of, 17
  clues to history of, 157–164
  composition of crust, 114
  density of, 112
  division into spheres, 111–127
  geological history of, 169–184
  inclination of axis, 97–99
  interior of, 111–115
  magnetic properties of, 113
  mass of, 91
  physical characteristics of, 89–93
  precession of axis, 96–97
  revolution about sun, 96
  rotation on axis, 93–95
  shape of, 89
  size of, 90–91 *
  speed in orbit, 96
  temperatures on, 92
Earth history,
  and fossils, 161–168
  and human history, 174
  bases of division, 169–178
  from rock correlations, 157–161
  principal events of, 176–184
  well-recorded, 174 *
Early man, 6–7
Earthquakes, 112, 144–148 *
Earthquake waves, 147–148
Eclipse,
  of moon, 106–107 *
  of sun, 105–106 *
Ecliptic, 97
Edison, Thomas, 13, 627, 628, 648
Edison Effect, 513–514
Edison storage battery, 465
Efficiency,
  of Diesel engine, 414 *
  of gas engine, 413, 414 *
  of simple machines, 378–380
  of steam engine, 403, 414 *
  of steam turbine, 404, 414 *
Egloff, Gustav, 347
Einstein, Albert, 13, 31, 391, 504
Electrical power, 473–474
Electrical resistance, 472, 473

Electric cells,
dry, 464
Edison, 465
lead storage, 465
primary, 464
voltaic, 461–464
Electric charges, 451–453
Electric circuit, 471–474
Electric current,
chemical effect of, 476–479
chemically generated, 460–465
conventional, 464, 466 *
electronic, 464
heating effect of, 474–475
magnetically generated, 466–470
mechanical effects of, 475–476
nature of, 459
related to static electricity, 459
Electric dynamo, 468–470
Electric eye, 470
Electric fish, 444
Electric furnace, 308
Electricity,
and the atom, 484–512
current, 459–481
flow through gases, 485–488
induced, 453–454
origin of word, 444
resinous and vitreous, 446
static, 443–456
Electric lighting, 648–649
Electric motor, 475–476
Electric refrigerator, 662–664
Electroluminescence, 650–653
Electrolysis, 476–478
Electrolytic dissociation, 462
Electromagnetic spectrum, 598–606
Electromotive force, 471–472
Electromotive series, 464
Electron,
and cathode rays, 491
and photoelectric effect, 494, 518
charge on, 488–493 *
discovery of, 488–491
e/m value, 488–491 *
mass of, 493–494
nature of, 449–450
Electronics, 513–534
Electronic vulcanization, 336 *
Electron microscope, 612–613 *
Electroplating, 478–479
Electroscope, 452 *
Electrostatic induction, 453–45
Electrotyping, 478, 626
Elementary magnet, 440
Elements,
alphabetical table of, 224–22
Aristotle's, 220 *
defined, 223
Dalton's, 222–223
periodic table of, 227
e/m for electron, 488–491 *
Energy,
and annihilation of matter, 31
definition of, 385
emitted by sun, 30–31
kinds, 385–390
kinetic, 388–390
potential, 386
Engines,
airplane, 412–413 *
airstream, 417, 418–422 *
continuous duct, 422 *
Diesel, 413–416 *
gas, 407–413 *
heat, 396–426 *
Hero's, 417 *
intermittent duct, 420–422 *
internal combustion, 405–416
Newcomen's, 398–399 *
ram-jet, 422 *
reaction, 416–424
rocket, 422–424
steam, 397–404 *
turbo-jet, 418–420 *
Equinoxes, 99
Eras, geological, 173–184
Eratosthenes, 17, 91
Erg, 375
Erosion, 133–141
Erosional mountains, 138
Ether, 594–595
Euclid, 7
Euler, 187
Evaporation, 232, 252
Exact sciences, 201

Exhaust manifold, 410
Expansion,
due to heating, 243–244, 254
of water, 267–268
Experimental facts, 2
Eye,
defects, 609 *
optical arrangement of, 608 *
optical aids for, 610–613

Fahrenheit temperature scale, 241
Faraday, Michael, 343, 467
Far-sightedness, 609 *
Fathom, 203
Faulting, 143–144 *
Federal Communications Commission, 600
Filament, 514
Filtration of water, 272
Fire damp, 291
Fire extinguisher, 266
Firefly, 653
Fission, 507
Fixed stars, 58
Fjords, 144
Flood plain, 137
Flotation, oil froth, 300
Flowering plants, 182
Fluorescent lighting, 652–653
Foam rubber, 336 *
Focusing,
of camera, 607
of eye, 608
Fog, 122
Folding, 143 *
Foot-candle, 638
Foot-pound, 375
Force, 369–374, 382
Ford, Henry, 13
Formulae, chemical, 228
Fossils,
abundance of, 163–164
and climate, 165–166
and earth history, 161–168
and sedimentary rock, 132
early ideas about, 164
kinds of, 161–163
Foucault, Leon, 94, 574
Foucault pendulum, 94
Four-stroke cycle, 410 *
Fractional distillation, 263–265
Franklin, Benjamin, 447–448
French, Sidney, 314
Frequency of waves, 540, 548, 550
Frictioning of rubber, 334
Frost, 122
Fuel injector, 415 *
Fuels,
gaseous, 290–293
liquid, 286–290
solid, 282–286
Fundamental tone, 562
Fundamental units, 202
Fuse, 474
Fusion, heat of, 657–658

Gage blocks, 206 *
Galactic nebulae, 59 *
Galaxies, exterior, 66–67
Galaxy, 52–61
Galilei, Galileo,
and atomic theory, 221
and first thermometer, 240 *
and Jupiter's moons, 19–20
and phases of Venus, 19
and speed of light, 573
and sunspots, 20
and the Inquisition, 20
and the telescope, 19, 71
importance to science, 10–11
Galle, 40
Galvani, Luigi, 460
Galvanized iron, 311
Gamma rays, 496, 605–606 *
Gas engine, 407–413 *
Gases, 232–235
Gas laws, 233–235, 253
Gasoline, 288–290
Gas refrigerator, 664–667 *
Gassendi, 221
Gay-Lussac law, 234, 253 *
Geissler, Heinrich, 485
Generator (see Dynamo)
Geocentric universe, 18
Geological,
cycles, 169–173 *

Geological — *continued*
eras, 173–184 *
history, 169–184
periods, 173 *
Geology, beginning of, 130–131
Geometry, 7
Gerbert, 197
German silver, 315
Giant stars, 57
Gilbert, Sir William, 429–430, 444
Glaciers, 139, 158–159
Glass, 351–352
Goodyear, Charles, 329
Gradation, 133
Grand Canyon, 136, 175 *
Granite, 131
Graphite, 284
Gravitation,
and mass of earth, 91
and mass of stars, 86
law of, 22, 370–371
Gray, Stephen, 446
Great Bear constellation, 62–64
Great Dog constellation, 65
Great Lakes, 144
Greeks,
and astronomy, 16–18
and calendar, 212
and nature of matter, 219–220
and number system, 192–193
and scientific method, 7
Greenland, 139
Gregorian calendar, 213–214
Grid, 515
Ground water, 138
Guericke, Otto von, 444–446
Gullies, 135 *
Gutenberg, John, 625

Haber, Fritz, 353
Habitability,
conditions essential for, 92–93
of planets, 36–37
Hail, 122
Hairy mammoth fossils, 162
"Half-life" of radioactive substances, 498
Hall, Charles Martin, 310
Halley, Edmund, 42
Halley's comet, 42–43
Handel, Georg Friedrich, 569
Hard water, 270, 272
Harmonics, 562
Harmonious tone combinations
Hawaiian Islands, 150
Heat engines, 396–426
Heat,
and expansion, 243–244, 254
and temperature, 239–243
caloric theory of, 246–249
conduction of, 244 *
convection of, 244–245
effect of absorption, 243, 253
mechanical equivalent of, 248
249, 390 *
methods of transfer, 244–246
nature of, 239–249
of fusion, 657–658
of vaporization, 658
radiated, 602–603
relation to matter, 243–246
units of, 243
Heat value of fuels, 283
Heavy hydrogen, 274
Heavy water, 273–274
Heligoland, 139
Heliocentric hypothesis, 19
Helium, 81, 116, 266
Helmholtz, Hermann von, 617
Hematite, 301
Henry, Joseph, 627
Hercules, 26
Herolt, Paul, 310
Hero of Alexandria, 417
Herschel, Sir William, 39
Hertz, Heinrich, 631
Hevea braziliensis, 325
High molecular weight compou
353–354
Himalayas, 143, 182
Hindu-Arabic numbers, 195–19
Hipparchus, 17, 55
Hippocrates, 8
Hogben, Lancelot, 71, 187, 188
Homer, 295
Homo sapiens, 183

Hooke's law, 374
Horse latitudes, 120, 122
Horsepower, 380–381
Hour, 209–210
Hudson River valley, 144
Hydrocarbons, 286–287
Hydrogen, 274–275
Hydrogenation, 275, 289
Hydrogen peroxide, 269
Hydrosphere, 115
Hypermetropia, 609 *

Ice ages, 179, 183
Iceland, 150–151
Ice manufacture, 661–662 *
Ice refrigerator, 659–660 *
Igneous intrusion, 148
Igneous rock, 131–132
Ignition temperature, 279, 280
Illuminants,
  comparative costs, 642–643
  history of use, 642, 644–647
Illumination,
  and glare, 640
  and shadows, 640
  comparative costs, 642–643
  incandescent sources, 643–650
  intensity considerations, 636–640
  sources of, 643–654
Imaginary numbers, 198
Incandescence,
  early sources, 643–647
  efficiency, 650
  electric sources, 648–650
Inch, 204 *
Inclination of earth's axis, 97–99
Inclined plane, 376, 377 *
Induced electric charges, 453–454
Inertia, law of, 371
Infra-red photography, 602–603 *
Infra-red rays, 602–603
Inquisition, 20
Insect fossils, 162
Insulators, 473
Intake manifold, 406
Intensity of sound,
  and amplitude, 548, 550
  in musical tones, 561
Interaction, law of, 373
Interference of light, 585–587, 617 *
Interferometer, 83
Intermittent duct engine, 420–422 *
Internal combustion engine, 405–416
International pitch, 568–569
Interval law, 564–565
Invar, 315
Inverse proportion, 233–234
Ionization, 461–462
Ionosphere, 119
Iron,
  in earth's crust, 114
  ores, 301
  reduction of ore, 302–305
  refining of, 305–308
Iron Age, 6
Isoclinic lines, 431, 432 *
Isogonic lines, 431 *
Isoprene, 343, 344
Isostasy, 142
Isostatic movements, 141–144
Isotopes, 502–503

Jeans, Sir James, 55, 154
Jewish calendar, 212
Johannson gage blocks, 206 *
Johns Hopkins University, 422
Joule, James Prescott, 248–249, 390
Joule of work, 375
Julian calendar, 212–213
Juno, 63, 64
Jupiter, 34, 35, 38
Jurassic period, 182

Kelvin, Lord, 201
Kennely-Heaviside layer, 119
Kepler, Johann, 10, 20–21, 370
Kepler's laws,
  and comets, 42
  statement of, 20–21
Kerosene, 290
Kew Botanical Gardens, 326
Kilocycle, 600
Kilowatt, 381
Kilowatt-hour, 474
Kinetic energy, 388–390
Kinetic theory, 250–256 *

Koroseal, 356
Krakatoa, 150
Krypton, 266

La Place, Pierre, 47
Laplacian hypothesis, 47–48
Latent heat, 236, 254, 657–658
Latex,
  direct use, 332, 338–339
  plantation treatment, 330–332
  sponge, 338
Laurentian Highlands, 179
Law,
  acceleration, 371–373
  Boyle's, 234, 253 *
  conservation of energy, 390–391, 392
  electric forces, 450
  definite composition, 228
  first thermodynamic, 390–391, 393
  Gay-Lussac's, 234, 253 *
  gravitation, 22, 370–371
  Hooke's, 374
  inertia, 371
  interaction, 373
  Kepler's, 20–21
  magnetic forces, 433
  musical intervals, 564–565
  planetary motion, 21
  reflection of light, 73, 577 *
  reflection of sound, 551
  variation of light intensity, 638 *
Lead, 312–313
Leap year rule, 213
Length,
  legal standard of, 205–206
  units of, 202–206
Lens telescope, 71–73
Lesser Bear constellation, 64
Leucippus, 220
Lever, 376
Levierrer, Urbain, 40
Life,
  and geological cycles, 171, 173
  essentials for maintenance of, 92–93
Light,
  and gamma rays, 605–606 *
  and infra-red rays, 602–603 *
  and ultra-violet rays, 604
  and x-rays, 604–605 *
  diffraction, 575–577
  Doppler effect, 587
  intensity, 637–638 *
  interference, 585–587 *
  medium for waves, 594–595
  nature of, 591–595
  polarization, 587–591 *
  reflection of, 73, 577–581 *
  refraction of, 581–585 *
  scattering, 618–619
  speed of, 572–575
  straight line propagation, 575
  visible range, 603
Lightning, 443, 447–448 *
Lightning rods, 455
Light sources, 644–647
Light year, 23
Lignite, 283
Limestone, 158, 161
Lines of force, 436, 437 *
Linotype, 625–626
Lippershey, Hans, 71
Liquid air, 263–264
Liquids, 232
Lithosphere, 113–115, 131–133
Little Dipper, 95 *
Lockyear, Sir Norman, 81
Lodestone, 427
Logarithms, 198
Long electromagnetic waves, 598–599
Long Island, 139
Longitudinal waves, 538
Lowell, Percival, 40
Lowell observatory, 40
Low-melting-point alloys, 314
Lucite, 358 *
Luckiesh, Matthew, 639
Luminol, 654
Lunar,
  craters, 100
  eclipse, 90, 106–107 *
  month, 103–104, 208

Lustron, 356
Lyra, 64

Machines,
and work, 376–380 *
simple, 376 *
Magma, 131
Magnalium, 316
Magnes, 427
Magnesia, 427
Magnetic,
alloys, 313
attraction and repulsion, 432–433, 437 *
declination, 431 *
dip, 432 *
elements, 440
field, 436, 437 *
poles, 430, 434, 437 *
saturation, 435, 439 *
Magnetism, 427–441
Magnetite, 301, 427
Magnetizing methods, 434, 437–439
Magnetostriction, 435–439
Magnitude of stars, 55–56
Major geological cycle, 170
Major triad, 566
Malleability, 296
Mammals, 164–165, 181, 182
Mammoth Cave, 138
Manhattan District, 509
Mantle rock, 133
Marble, 132
Marconi, Guglielmo, 13, 631
Mars, 34, 35, 37–38 *
Marsh gas, 282, 291
Mass-energy equivalence, 504
Mass, legal standard of, 208
Mass units, 206–208
Matter,
kinds of, 222–231
kinetic theory of (see Kinetic theory)
nature of, 218–257
relation to heat, 243–246
states of, 231–235
Mauna Loa, 149
Maxwell, James Clerk, 630
Mayan numbers, 193–194
Meanders, 137
Measurement,
derived units of, 215
fundamental units of, 202–210
of speed of light, 573–575 *
role in science, 201–202
Mechanical advantage, 377–378
Mechanical equivalent of heat, 248–249, 390 *
Mendeleef, D. L., 226
Mercury, 34, 35, 36
Mercury-vapor lamp, 651–652
Mergenthaler, Otto, 625
Mesozoic era, 176, 181–182
Messier nebula, 67 *
Metallic luster, 296
Metallurgical processes, 299–300
Metals,
general properties, 296–299
historical background, 295–296
metallurgical processes, 299–301
Metamorphic rock, 132–133
Meteor Crater, 45
Meteorites, 44, 112–113 *
Meteorological balloons, 126–127
Meteor showers, 45
Meteors, 43–45
Meter, 205 *
Methane, 282, 287, 291, 344
Metric system, 205
Michelson, Albert, 83, 575
Micron, 320
Microscope,
electron, 612–613 *
invention of, 71
optical, 611
ultra-, 322
Middle Ages and astronomy, 18
Middle Ages and science, 9–10
Midnight sun, 100 *
Milky Way, 54 *
Miller, Dayton C., 564
Millikan, R. A., 13, 491
Mississippian period, 181
Mississippi River, 135, 141 *
Mixtures, 230–231

Molecular theory of magnetism, 435–440
Molecules, 288
Monatomic gases, 229
Monel metal, 316
Month, lunar, 103–104
Month names, 213
Moon,
  and atmosphere, 100–101
  craters on, 100
  motions of, 102
  phases of, 104–105 *
  radar contact with, 528 *
  size of, 100
  temperatures on, 101–102
Morse, Samuel B., 626
Motion, Newton's laws of, 371–373
Motor, electric, 475–476
Moulton, F. R., 48
Mountains,
  dome, 149
  erosional, 138
Mount Everest, 114
Mount Palomar telescope, 74, 78 *
Mount Wilson telescope, 77
Movable type, 625
Multiplication of Roman numbers, 197–198
Münters, Carl, 665
Musical interval, 564–566
Musical scales,
  diatonic, 565–567
  tempered, 567–568
Music, physical basis of, 561–569
Myopia, 609

Napoleon III, 310
National Defense Research Committee, 509
Natural gas, 282, 291
Nature,
  of charged and uncharged bodies, 451–452
  of color, 613–616
  of combustion, 279
  of electricity, 448–450
  of light, 591–595
Nature — *continued*
  of magnetism, 435–440
  of matter, 218–257
Neap tides, 108
Near-sightedness, 609 *
Nebulae,
  extra-galactic, 66–67
  galactic, 59–61
Nebular hypothesis, 47–48
Negative electricity, 446, 448–449
Neolithic culture, 6
Neolithic man, 183
Neon, 116, 266
Neon lights, 650–651
Neon tube, 486
Neoprene, 345
Neptune, 34, 35, 40
Neutron, 499, 502, 507
Newcomen, Thomas, 398
Newlands, 226
Newton, Sir Isaac,
  and color, 613–615 *
  and law of gravitation, 21
  and nature of light, 3, 592–593
  and reflecting telescope, 73
  and white light, 78
  importance to science, 11
Newton's laws of motion, 369–373
Nieuwland, Julius, 346
Nitrogen, 116, 265
North America in Paleozoic era, 178 *
North-seeking pole, 430
North star, 64
Novae, 57
Nuclear atom, 499–503
Number names, 189–190
Numbers and number systems, 187–198
Number system bases, 190–191
Number systems,
  Greek, 192–193
  Hindu-Arabic, 195–198
  Mayan, 193–194
  Roman, 192
Number writing, 191–192
Nylon, 362

Obliquity of the ecliptic, 97
Occultation, 107
Ocean temperatures, 115
Oersted, Hans Christian, 466
Ohio hills, 138
Ohm, 472, 473
Oil drop experiment, 491–493 *
Oil froth flotation, 300
Olcott, W. T., 62
Open hearth furnace, 307 *
Opera glasses, 610
Optical instruments, 610–613
Orbits,
  of comets, 42
  of planets, 21, 23
Ordovician period, 180
Organic compounds, 350
Organic synthetics,
  drugs, 364–365
  dyes, 364
  general characteristics, 353–354
  plastics, 353–367
  raw materials for, 360–361
Ounce, 207
Overtones, 562–564 *
Oxidation, 265
Oxyacetylene torch, 293
Oxyacetylene welding, 265, 281
Oxygen,
  in air, 116
  in earth's crust, 114
  nature and properties, 264–265
Ozark Mountains, 179

P-80 Plane (U.S.A.A.F.), 421 *
Pale crepe rubber, 331
Paleozoic era, 176, 180–181, 282 *
Palisades of the Hudson River, 148
Palm, 203
Panning for gold, 299
Pantheon, 94
Parallax of stars, 81–82 *
Parsons, Charles, 403
Particle theories of light, 592–593
Peary, Admiral, 44
Peat, 283
Pennsylvanian period, 181
Pentadecane, 287
Pentatriacontane, 287
Pentode, 517–518
Periodic law, 226
Periodic table of elements, 227
Period of waves, 542, 549, 550 *
Perkin, William, 364
Permalloy, 313, 317, 441
Permeability, 440
Permian period, 181
Petrified wood, 162
Petroleum, 287–288
Pewter, 316
Phases,
  of Mercury, 36
  of moon, 104–105 *
  of Venus, 19-36 *
Photoelectric cell, 470, 518–519
Photoelectric effect, 470, 494 *
Photoflash lamps, 649
Photoflood lamps, 649
Photosphere, 26
Physical basis of music, 560–569
Piazzi, 41
Picture writing, 191
Pig iron, 304, 305
Pin-hole camera, 576 *
Pitch,
  and frequency, 548, 550
  and wave-length, 548–549, 550
  of musical tones, 561
  types, 566–569
Plan Position Indicator (PPI), 524–525 *
Plane polarized waves, 588
Planetary motion, laws, 21 *
Planetesimal hypothesis, 48–49
Planetoids, 41–42
Planets,
  and habitability, 92–93
  distance to, 81–83
  general description, 34–41
  mass of, 86
  measurements made on, 78–86
  size of, 83–84
  table of data about, 35
Plants and erosion, 141
Plastics (see Synthetic plastics)
Platen, Baltzar von, 665

Plato, 8, 592
Pliny, 351
Pliofilm, 342, 363
Pluto, 34, 35, 40–41
Polar Atlantic air, 126
Polar continental air, 126
Polaris, 64, 94–95, 96
Polarization of light, 587–591 *
Polar Pacific air, 126
Polaroid glass, 589–591 *
Pole star (see Polaris)
Polonium, 495
Polymer, 344
Polymerization, 356
Polythene, 356
Pope Gregory XIII, 213
Positive electricity, 446, 448, 449
Positron, 449
Potential energy, 386–387 *
Pound, 207, 208, 373
Poundal, 374
Power, 380–382
Precession of earth's axis, 96–97
Precipitation and weather, 121–122
Prehistoric earth history, 174
Pressure,
  atmospheric, 260–261
  in gases, 233–234, 253
Priestley, Joseph, 264, 265, 352
Primates, 182
Printing, 624–626
Printing telegraph, 632–633
Producer gas, 292
Proterozoic era, 176, 179
Proton, 449
Proximity fuze, 529–532 *
Ptolemaic universe, 18 *
Ptolemy, Claudius, 9, 18
Puddling furnace, 305
Pulley, 376
Purposes of science, 11–13
Pyrenees, 182
Pyrex, 352
Pythagoras, 7–8, 556

Quality of musical tones, 526–528
Quaternary period, 183
Quinine, 364

Radar, 520–529 *
Radiation of heat, 246
Radio,
  and vacuum tubes, 513–518
  broadcast bands, 598–601
  early development, 630–632
  ultra-short waves, 601–602
Radioactive disintegration, 156
Radioactive elements, 498
Radioactivity,
  and age of the earth, 91–92
  discovery of, 494–495
  nature of, 495–498
  table of elements, 498
Radio picture transmission, 633
Radio receiving circuit, 516 *
Radio tubes (see Vacuum tubes)
Radium, 131, 156, 495, 498
Radon, 266
Rain, 121
Rainbow, 80 *
Rainfall, 134
Ram-jet engine, 422 *
Ramsay, Sir William, 81
Rarefaction, 546
Rare gases, 266
Rayon, 361
Reaction engine, 416–424 *
Receiver, telephone, 628–629 *
Reciprocating motion, 401
Reclaimed rubber, 333–334
Rectification, 515
Reduction,
  defined, 275
  of metallic ores, 300
Refining of metals, 300–301
Reflecting telescope, 73–77
Reflection,
  of light, 73, 577–580 *
  of sound, 549–555 *
  total, 584 *
Reflectors of light, 577–581 *
Refracting telescope, 71–73 *
Refraction,
  distinguished from diffraction, 577 *
  of light, 581–585 *
  of sound, 555–556 *

Refrigeration, 271, 657–667
Refrigerators,
compressor-type, 660–664 *
gas, 664–667 *
ice, 659–660 *
Region of auroral displays, 119
Region of meteors, 118–119
Relative humidity, 121
Relief-building factors, 141–151
Relief-levelling factors, 133–141
Reptiles, 181, 182
Resins (see Synthetic plastics)
Resistance, electrical, 472
Retentivity, 440
Retina of eye, 608
Retrograde satellites, 46
Reverberation, 551
Reversing layer of sun, 26
Revolution, geological, 169, 171
Revolution of earth, 96
" Roaring Forties," 120
Roasting, metallurgical, 300
Rock correlations,
and earth history, 157–161
limitations of, 161
Rock formations, 157–160, 164
Rock of Gibraltar, 133
Rock, types of, 131–133
Rockets, 417, 422–424
Rocky Mountains, 131, 179, 182
Rod, early definition, 204 *
Roman calendar, 212
Roman numbers, 192, 197–198
Romer, Olaf, 573
Rotation of earth, 93–95, 120
Rotation of sun, 20, 25–26
Rubber,
and World War I, 327
application, 337–341
bearings, 339
chemical properties, 328–329
compounding ingredients, 332–334
derivatives, 342
distribution U. S. consumption, 337
formula for molecule, 328, 344
kinds of crude, 331, 332
nature and properties, 328–330
Rubber — *continued*
new uses, 339–341
plantation procedure, 330–332 *
physical properties, 329–330
sources, 325–328
steps in processing, 330–337
structural formula, 350
synthetic, 342–348
U. S. importation, 326–327
vulcanization, 333, 334–337
world statistics, 326
Rubber Reserve Corporation, 327, 345
Rumford, Count (Benjamin Thompson), 247, 390
Rusting, 297
Rutherford, Sir Ernest, 505

Sabre-toothed tiger fossils, 162, 165
Safety glass, 357 *
Saint Elmo's fire, 444, 454
San Francisco earthquake, 145, 147
Saran, 356
Saros, 106, 209
Saturation,
atmospheric, 121
magnetic, 436
Saturn, 34, 35, 38–39 *
Savery, Thomas, 398
Scales, musical, 565–569
Scattering of light, 618–619
Schade, James W., 339
Science,
and the scientific method, 1–15
definition of, 1, 5
derivation of word, 1
in Middle Ages, 9–10
purposes of, 11–13
Scientific attitude, 4–5
Scientific laws, 12–13
Scientific method,
analysis of, 2–5
and early church, 9–10
examples of application, 2–4
historical development of, 5–11
steps in, 2–3
Scientific theories, 12–13
Screen grid, 517

Screw, 376 *
Seasons, 98 *
Second of time, 210
Sedimentary rock, 132
Seegar's cones, 241
Seismograph, 147 *
Seismology, 145
Seven Metals, The, 296
Shale formation, 158
Short-wave radio, 601
Siberian meteorite, 45
Sierra Nevada Mountains, 143, 182
Sight,
  color vision, 617
  physical basis of, 606–613
Silicon, 114
Silicones, 363–364
Sill, 148
Silurian period, 180
Simple machines, 376 *
Sirius, 56, 65, 209
Slag, 305
Slate, 132
Sleet, 121
Slide-valve steam engine, 402–403 *
Smoked-sheet rubber, 332
Smoke elimination, 324 *
Smyth, H. D., 503
Snow, 122
Socrates, 428
Sodium vapor lamps, 651 *
Soft water, 272
Soil, 133
Solar eclipse, 105–106 *
Solar energy, 30–31
Solar prominence, 27 *
Solar system,
  Copernican model, 19
  diagram of, 24 *
  general picture of, 22–24
  location in galaxy, 52–53
  members of, 25–45
  origin of, 45–49
  planets in, 23
  Ptolemaic model, 17 *
  scale model of, 34–36
  speed of, 24
  universe beyond, 52–69
Solar year, 208
Solder, 316
Solenoid, 467 *
Solids, 231–232
Solution pressure, 462
Solutions, 231
Sound,
  beats, 557–558 *
  Doppler effect, 559–560 *
  essentials for producing, 54
  kind of waves, 545–546
  properties of waves, 547–5
  reflection, 549–555 *
  refraction, 555–556 *
  speed in various media, 547
Sounding balloon, 126–127
South-seeking magnetic pole,
Span, 203
Spark plug, 409
Spectacles, 70
Spectral lines,
  and distance to stars, 82–83
  and speed of stars, 85
  and water on moon, 101–10
Spectroscope, 80
Spectrum,
  absorption, 81
  and star velocities, 85
  bright-line, 80
  continuous, 80
  dark-line, 81
  electromagnetic, 598–606
  of stars, 81
Speech, 623
Speed,
  and acceleration, 371–372
  of light, 572–575 *
  of sound, 547–548, 550
Spontaneous combustion, 279–
Spring tides, 108
Stainless steel, 317
Standard candle, 636–637
Standard pitch, 568–569
Star charts (see inside covers)
  directions for using, 61–62
Stars,
  and constellations, 61–66
  colors of, 57, 84

Stars — *continued*
composition of, 78–81
distances to, 81–83
dwarfs and giants, 58
first magnitude table, 56
fixed, 58
life cycle of, 57–58
magnitude definition, 55–56
mass of, 86
motion of, 58
number in our galaxy, 55
sizes of, 83–84
speed of, 85
temperatures of, 57, 84–85
variable, 57
Star trails, 94 *
Static discharges, 454–456
Static electricity, 443–456
Steam engines, 397–403 *
Steam turbine, 403–405 *
Steel, 306–308
Steel alloys, 314, 316
Steinmetz, Charles, 13
Stereotyping, 626
Sterling silver, 316
Stomach stones, 163
Storage battery, 465
Stratification of atmosphere, 116–119 *
Stratosphere, 117-118 *
Structural formulae, 344
Submergence, geological, 169, 170
Sulfa drugs, 364
Sun,
and weather, 119
as the source of energy, 30, 31, 389–390
physical characteristics of, 25–33
Sunspot cycles, 28–30
Sunspots, 20, 28–30
Super-cooled liquid, 329
Suspension, colloidal, 319–325
Swann, W. F. G., 46, 251
Symbols, chemical, 223
Synthetic drugs, 364–365
Synthetic dyes, 364
Synthetic fibers, 361–362
Synthetic plastics,
acryloids, 358–359
definition, 355
raw materials for, 360–361
thermoplastic, 357
thermosetting, 359
types, 355–360
vinyls, 356–358
Synthetic rubber,
and World War I, 343, 345–346
cost, 347
historical background, 342–346
production outlook, 347–348
properties and applications, 346
Synthetics,
definition, 350
inorganic, 351–353
organic, 353–367
Synthetic soap, 365

Taconic Mountains, 180
Telegraph, 626–627
Telephone, 627–630 *
Telescope,
amateur-built, 76 *
and Galileo, 19
examples of, 77–78 *
historical background, 70–71
light-gathering power of, 72–74
reflecting, 73–77 *
refracting, 71–73 *
Teletype, 632
Television, 633
Temperature,
and heat, 239–243
effect on combustion, 279–280
Temperature measurement,
and thermocouple, 84–85
relation to color, 84
Temperature scales,
absolute, 254–255 *
centigrade, 241–242 *
comparison of various scales, 256 *
conversion between scales, 242
Fahrenheit scale, 241 *
Tempered scale, 567–568
Terminal moraine, 139
Tertiary period, 182

Tetrode, 517–518
Thales of Miletus, 7
Thermal-jet engine (see Turbo-jet engine)
Thermocouple, 84, 240–241, 470
Thermodynamics,
  first law of, 390–391, 393
  second law of, 392–393
Thermoelectric effect, 470
Thermometers,
  Galileo's, 240 *
  modern types, 240–243
Thermostat, 243 *
Thomson, Sir J. J., 502
Thompson, Benjamin (Count Rumford), 247
Tides, 107–108 *
Timbre, 562–564
Time units, 208–210
Tin, 312
Tin foil, 316
Tires and tubes, 337–338
Tone, fundamental, 562
Torricelli, Evangelista, 261
Total reflection, 583, 584, 610 *
Trade winds, 120
Transfer of heat, 244–246
Transformer, 480–481
Transit, 107
Transmutation, artificial, 505
Transverse waves, 537–538, 589 *
Tree rings, 28, 30
Triangulation, 81 *
Triassic period, 181
Trilobites, 165, 180
Triode, 515-517
Tropical Gulf air, 126
Tropopause, 117
Troposphere, 117 *
Trough of wave, 537
Tubing of rubber, 334
Tudor, Francis, 271
Turbine, steam, 403–405 *
Turbo-jet engine, 418–420 *
Tuscarora Deep, 114
Twain, Mark, 122
Tyndall effect, 323 *
Type metal, 314, 625
Typhoid fever, 272

U. S. Army Signal Corps, 528
Ultra-microscope, 322 *
Ultra-violet light, 604
Unconformity, 158 *
Units,
  based on water, 273
  derived, 215
  fundamental, 202
  of atomic weight, 223–225
  of electricity, 472, 473, 474
  of force, 373–374
  of heat, 243
  of length, 202–206
  of light, 636–638
  of mass, 206–208
  of power, 380–381
  of time, 208–210
  of work, 375
Universe beyond the solar systen 69
Unpolarized waves, 588
Uranium, 131, 156, 507–510
Uranium X, 156
Uranus, 34, 35, 39–40
Ursa Major, 62–64
Ussher, Archbishop, 155

V-1 bomb (German), 421, 526
V-2 bomb (German), 423
Vacuum tubes, 513–518 *
Valley glacier, 139
Valleys, history of, 136–137
Van de Graaff generator, 446 *
Vaporization, heat of, 658
Variable stars, 57
Vega, 64–65, 96
Vegetable cooking fats, 275
Venus, 34, 35, 36–37 *
Vinylite, 356
Vinyl polymers, 356–358
Visible light, 603
Vision, defects in, 609
Vitamin D, 604
Volcanic action, 148–151
Volcanic belts on earth, 146 *
Volcanic plug, 151
Volcanoes, 149–151

Volt, 471–473
Volta, Allessandro, 460
Voltaic cell, 461–464 *
Voltaic pile, 460–461
VT fuze (see Proximity fuze)
Vulcanism, 148–151 *
Vulcanization,
  of rubber, 333, 334–337
  of synthetic elastics, 348

Water,
  and disease, 272
  and gradation, 134–139
  as a measuring standard, 273
  as a solvent, 269
  composition, 269–271
  expansion properties, 267–268 *
  for drinking, 271–272
  hard and soft, 270, 272–273
  heavy, 273–274
  impurities in, 269–270
  latent heats of, 657–658
  occurrence, 266–267
  properties, 267–269
  purification, 272
  uses, 271–273
Water-borne diseases, 272
Water gas, 291–292
Water table, 138
Watt, 381, 473
Watt, James, 399
Wave-length,
  definition, 540 *
  of green light, 550, 587
  of red light, 587
  of sound, 548–549, 550
  of violet light, 587
Wave motion, 535–542
Wave theory of light, 593–594
Waves,
  amplitude of, 541
  and erosion, 138
  diagram of, 539 *
  essentials for, 536–537
  frequency of, 540
  kinds of, 537–539
  length, 540
  period, 542
  sound, 545–546 *
Weather, 119–127
Weather Bureau, United States, 123
Weather forecasting, 122–127
Weather map, 123 *
Wedge, 376
Week, 209
Well-recorded earth history, 174 *
Westerly wind belts, 120
Wheel and axle, 376
" Whispering gallery " effect, 555
Wickham, Henry A., 326
William the Conqueror, 42
Wilson cloud chamber, 496–497
Wind and erosion, 140–141 *
Wind belts, 120 *
Woods metal, 314–315
Woolworth Tower, 154
Wooly rhinoceros, 162
Work, 249, 374–380, 382
World War II,
  and radar, 526–529
  and proximity fuze, 529–532
  and rubber, 327
  and synthetic rubber, 347
Worm burrows, 163
Wright, Orville, 13
Writing, 623
Wrought iron, 305–306

Xenon, 266
X-rays, 604–605 *

Yard,
  early definition, 203, 204
  legal definition, 206
  relation to meter, 206 *
Year, 96
Yerkes telescope, 73, 77 *
Young-Helmholtz theory, 617
Young, Thomas, 585, 593, 617

Zero,
  absolute, 255
  in Hindu-Arabic system, 195
  in Mayan numbers, 193
  uses of, 196–197
Zinc, 311–312

ALLAMA IQBAL LIBRARY
22765

JAMMU & KASHMIR UNIVERSITY
LIBRARY
No 22765
Date 12.5.58
SRINAGAR